

Vida Zsófia Viktória

Távolság a tudományos együttműködési hálózatokban

TANULMÁNYOK A TUDOMÁNYELEMZÉS MAI GYAKORLATÁBÓL

1

SOROZATSZERKESZTŐ:

SOÓS SÁNDOR

Vida Zsófia Viktória

TÁVOLSÁG A TUDOMÁNYOS EGYÜTTMŰKÖDÉSI HÁLÓZATOKBAN



Budapest, 2020

ISBN 978-963-7451-55-3

DOI 10.36820/tudomanyelemzes.2020.1

ISSN 2677-1683

A kötet a szerző

Távolság a tudományos együttműködési hálózatokban

című PhD értekezésének felhasználásával készült.

ELTE TTK Földtudományi Doktori Iskola

Földrajz–Meteorológia Program

témavezető: Jakobi Ákos PhD, habil.

Budapest, 2018

Felelős kiadó: az MTA Könyvtár és Információs Központ főigazgatója

A sorozat szerkesztője: Soós Sándor

A borítón szereplő kép forrása: Freepik.com

Tipográfia és tördelés: Vas Viktória

Nyomta és kötötte az Alföldi Nyomda Zrt., Debrecen

Felelős vezető: György Géza vezérigazgató

TARTALOMJEGYZÉK

Előszó	7
Zsófia Viktória Vida: Distance on research collaboration networks	9
1. Bevezetés	11
1.1. Témafelvetés, a téma aktualitása	11
1.2. Célok	12
1.3. Az értekezés felépítése	14
2. Bevezetés a hálózattudományba	16
2.1. A hálózatok fogalmi alapjai	16
2.2. A hálózattudomány fő mérföldkövei	18
3. Tudományos együttműködési hálózatok	29
3.1. Tudományos együttműködések	29
3.1.1. Mit értünk tudományos együttműködés alatt?	29
3.1.2. Tudományos együttműködés és társszerzőség viszonya	31
3.1.3. Miért jönnek létre tudományos együttműködések?	33
3.2. A kutatói együttműködések empirikus vizsgálata	38
3.2.1. Az együttműködések hálózatelemzési módszerekkel vizsgáló tanulmányok	39
3.2.2. Az együttműködések részben hálózatelemzési módszerekkel vizsgáló tanulmányok	43
4. Távolság értelmezése tudományos kapcsolathálózatokon	47
4.1. Térkategóriák értelmezése hálózatokon	47
4.2. Távolságtípusok a tudományos együttműködésekben	54
4.2.1. Földrajzi távolság	55
4.2.2. Társadalmi távolság	57
4.2.3. Kognitív távolság	61
4.3. Társadalmi és kognitív távolság viszonya a szerzői bibliográfiái csatoláson keresztül – egy konceptuális modell	71
5. Bibliometriai adatok elérése, kezelése – a vizsgálat adatbázisa	76
5.1. Az „ideális” adatbázis	76
5.2. A tudományometriában leggyakrabban használt adatbázisok rövid bemutatása	77
5.3. A kutatásba vont adatok	80
5.4. Az adatok felhasználásának korlátai	86

5.4.1. A WoS adatbázisból adódó torzítások	87
5.4.2. Az adatok előkészítése a vizsgálat lefolytatásához	91
6. Távolságtípusok meghatározása	97
6.1. Társadalmi távolság meghatározása	97
6.2. Kognitív távolság meghatározása	99
6.2.1. Teljes kognitív távolság meghatározása	100
6.2.2. Tiszta kognitív távolság meghatározása	101
6.3. Földrajzi távolság vizsgálatának módja	102
7. A létrehozott távolságtípusok összehasonlítása	108
7.1. Távolságtípusok összehasonlítása hálózatifüggetlen módszerekkel	109
7.1.1. A társadalmi és a kognitív hasonlóságmátrixok kapcsolata	109
7.1.2. Hierarchikus klaszteranalízis által létrejött tématerületek összehasonlítása	110
7.2. Távolságtípusok összehasonlítása hálózatelemzési módszerekkel	118
7.2.1. A társadalmi és a kognitív hálózatok jellemzése	119
7.2.2. A társadalmi, a teljes és tiszta kognitív hálózatok összehasonlítása QAP korreláció segítségével	125
7.2.3. Szerzői csoportok által kijelölt tématerületek az egyes távolságmátrixokban hálózati modularitással	128
7.3. A társadalmi, teljes és tiszta kognitív távolságtípusok a földrajzi térben	136
7.4. Az empirikus eredmények összefoglalása	144
8. Összefoglalás	147
Felhasznált irodalom	152
Internetes források	169
Ábrák jegyzéke	170
Táblázatok jegyzéke	172

Előszó

Kedves Olvasó!

Nagyon örülök, hogy a kezeden tartod ezt a könyvet, mert ez azt jelenti, hogy a könyvformában megjelent doktori disszertációm valami miatt felkeltette az érdeklődésedet.

Remélem, sikerült egy olyan munkát alkotnom, melyben olyan elgondolkodtató, számodra fontos kérdésköröket is találsz, amik tovább visznek Téged a saját munkád során. Engedd meg, hogy ebben a rövid előszóban az értekezés hivatalos nyelvezete előtt egy sokkal személyesebb hangvételben szóljak. Szeretnék köszönetet mondani azoknak, akik nélkül nem jöhetett volna létre ez a disszertáció, illetve egy kis bepillantást nyújtani a mű keletkezésének fő fázisaiba.

Először is szeretném megköszönni szűkebb és tágabb családomnak a támogatást, ami lehetőséget biztosított számomra, hogy az írásra koncentráljak. Csabival, férjemmel mindig meg tudtam beszélni a hirtelen felmerült kétségeimet: már az, hogy megfogalmaztam a nehézséget, segített átlendülni rajta. A nyelvi lektorom is ő volt egy-egy rész elkészültekor. Kisfiam nagyon türelmes baba volt és hagyta Anyát dolgozni. Édesanyám és anyósom felváltva sétáltatták a kisfiamat, boldogan játszottak vele, és elláttak minket finom ételekkel, hogy én nyugodtan írhattak. Az ő támogatásuk nélkül a cél előtt buktam volna el.

Azonban a disszertáció keletkezése sokkal régebbre nyúlik vissza, egészen oda, amikor még azt sem tudtam, hogy én valaha fogok egyáltalán doktori értekezést írni. Az egyetemen szabadon választható tárgyként felvettem a „Hálózatok világában” című kurzust. Ez az óra és az oktató, Gulyás László volt az, akinek hatására érdeklődésem a hálózatok felé fordult. Szakdolgozatomhoz még nem mertem hálózattudományi témát választani, mivel még magam is akkor ismerkedtem ezzel az azóta már egyre szélesebb körben ismertté vált területtel. Azonban mikor eldöntöttem, hogy jelentkezem doktori képzésre, már kifejezetten a hálózatokkal szerettem volna foglalkozni. A doktori képzésre való jelentkezés előtt és után is sok segítséget, támogatást kaptam Gulyás Lászlótól, aki szakirodalmat ajánlott és bátorított engem.

A következő fontos állomást a doktorandusz éveim jelentették. Nagy örömmre sikerült a felvételem az ELTE TTK Földtudományi Doktori Iskolájába, ahol három csodálatos doktori évet töltöttem, és a kutatás mellett az oktatásban is bekapcsolódhattam, ami nagyon inspirált engem. Témavezetőm Jakobi Ákos volt, akinek ezúton is köszönöm, hogy végigkísért ezen a hosszú úton, amíg egy tématerület iránti érdeklődésből kialakult egy kutatási kérdés és

végül megszületett a disszertáció. Sokat beszélgettünk, így tanúja volt annak, ahogy a szélesebb körű érdeklődésem egyre fókuszáltabbá vált – természetesen néhány zsákutcával és éles kanyarral tarkítva. Köszönöm Nemes-Nagy József tanár úrnak mindig tömör és bölcs tanácsait, valamint azt, hogy azt éreztem, mindig hisz bennem. Örülök, hogy Lócsei Hajnalka volt a szobatársam a doktori éveim alatt, akinél jobb kollégát nem is kívánhatott az ember.

A doktori kutatás során volt egy pillanat, amikor bár azt tudtam, hogy a hálózattudomány és a regionális tudomány elmélete érdekel, nagy kihívás elé állított, hogy ezt a két területet hogyan tudom felfűzni egy gondolatmenetre, hozzáférhető adatokkal alátámasztva. Egy tanszéki beszélgetés során Szabó Pál érdeklődéssel hallgatta végig a dilemmáimat, és felhívta a figyelmemet a kutatók és a hivatkozási adatok lehetőségére, melyek jól dokumentáltan elérhetők és földrajzi tartalommal is rendelkeznek. Ezért végtelenül hálás vagyok neki, mert így kezdtem bele több, ilyen adatokat feldolgozó kisebb tanulmányba, ami nagyban hozzájárult ahhoz, hogy a doktori évek után sikerült elnyernem egy, a témához kapcsolódó fiatal kutatói állást.

A doktori évek után a disszertáció megírása szempontjából végső és legfontosabb állomás az MTA Könyvtár és Információs Központ Tudománypolitikai és Tudományelemzési Osztályán meghirdetett fiatal kutatói pályázat elnyerése volt. Itt egy támogató és jól működő közösségbe kerültem. Közelebbről megismertem a tudományometriát, és elmélyedtem annak elemző tudományterképező irányzatában. Ez segített hozzá a felismeréshez, hogy a kutatói közösségek vizsgálata révén lehetőségem adódik a regionális tudomány távolsággal foglalkozó elméletének és a hálózattudomány módszereinek az ötvözésére.

Azért is köszönettel tartozom az MTA Könyvtár és Információs Központjának, hogy doktori kutatásom a jelen kötetben megjelenhetett. Vas Viktóriának külön köszönöm a gondos szerkesztést.

Hálás köszönetet szeretnék mondani végül, de nem utolsósorban Soós Sándornak, aki szakmai mentorom lett, és akihez a mindennapos munkához kapcsolódó kérdések mellett a disszertációmhoz kötődő problémákkal is fordulhattam. Az értekezés végleges koncepcióját is először vele osztottam meg.

Köszönöm, hogy elolvastad a disszertáció létrejöttének ezt a rövid, személyes történetét, mielőtt magát a kutatást is megismered. Remélem, tetszeni fog. Amennyiben kérdésed, további javaslatod lenne, bátran tedd meg a vida.zsolia@konyvtar.mta.hu e-mail címen.

Budapest, 2020. március 26.

Vida Zsófia Viktória

Zsófia Viktória Vida

Distance on research collaboration networks

Research collaboration is a widely analysed topic since the 1960s. Research collaborations form networks between researchers, which can be investigated with network analysis methods. Numerous factors influence the evolution of research collaborations. Along these factors we can measure the distance between collaborators. In general, the smaller is the distance between actors the more likely they will engage in collaboration.

The aim of the thesis is to determine and describe different distance-types between collaborators and to define their measurement possibilities. My research questions are:

1. Which types of distance can be defined in research collaboration networks?
2. How can cognitive distance be projected to the level of authors?
3. How can we interpret full cognitive distance?
4. How the different distance-types are related to each other?
5. Is there any difference between patterns of different science fields, namely Economics and Physical Geography?

In the empirical part of the thesis I investigated WoS records in Physical Geography and Economics between 2010–2014, which had at least one author with a Hungarian affiliation. As a result, I separated three distance-types: geographical, social and cognitive distance. In determining cognitive distance, during the extension of the bibliographic coupling analysis method to the author level, the references of one paper were assigned to all authors of the paper. Thus, there are two possibilities for a pair of authors to have a common reference: a) the two authors are co-authors, so both of them get the common reference, b) the two authors are not co-authors but they use the same reference because of their common research interest. I separated these two cases: the first one is the *social component* that can be determined with co-authorship networks, the second one is the *pure cognitive component*, which can predict potential collaborations. The sum of the two cases is the *full cognitive distance*, given by the results of author-level bibliographic coupling.

Comparing the two components with network-free and network analysis procedures, the dominance of the social factor was found in the full cognitive distance. The network structures defined by pure cognitive and full cognitive distance were found to be similar. While analyzing the strength of links the social distance and the full cognitive distance have shown more similarity.

The more frequent collaboration caused a stronger social component, so the full cognitive distance between authors in Physical Geography was smaller, whereas in the case of pure cognitive distance, I found stronger similarity between authors in Economics.

1. Bevezetés

1.1. Témafelvetés, a téma aktualitása

A tudományos együttműködések mindig is a tudás létrehozásának fontos színterei voltak. A kutatói együttműködések tanulmányozása egészen az 1960-as évekre nyúlik vissza (Glänzel és Schubert, 2004), és napjainkban is széles körben kutatott területet képvisel (Melin és Persson, 1996; Sonnenwald, 2007). Jelentőségét az is mutatja, hogy a társszerzős publikációk aránya folyamatosan növekvő tendenciát mutat (Smith, 1958; Clarke, 1964; de Solla Price és Beaver, 1966; Katz, 1994; Katz és Martin, 1997), emellett az egyes cikkek létrehozásában közreműködők átlagos száma is emelkedik, ahogy azt például Glänzel és Schubert (2004) kimutatta.

Ebből kiindulva állíthatjuk, hogy a kutatói együttműködések egyre nagyobb jelentőséggel bírnak, éppen ezért megismerésük kiemelt figyelmet érdemel. A szakirodalomból egyértelműen kiderül, hogy az együttműködések nem véletlenszerűen szerveződnek, létrejöttükben számos tényező játszik közre. Értekezésemben az egyik legfontosabb tényezőt, a *távolságot* fogom vizsgálni. A kutatói együttműködések egyik „kézzelfogható” megjelenési formáján, a társszerzőségeken keresztül mutatom be, hogy a távolság – és ennek számos típusa – miként hat a kapcsolatok létrejöttére és intenzitására.

A tudományos együttműködések létrejöttét befolyásoló tényezők egy-egy dimenzióként is felfoghatók. Az egyes dimenziók mentén mérhetővé válik, hogy két szereplő között mekkora a távolság. Minél kisebb távolságot találunk a két szereplő között, annál inkább hasonlítanak egymáshoz, és ez az esetek zömében segíti az együttműködések kialakulását (lásd pl. Paier és Scherngell, 2011). Fontos azonban kiemelni már itt is, hogy a nulla távolság, azaz a teljes hasonlóság gátló tényezőként is megjelenhet az együttműködésekben (lásd pl. Wuyts et al., 2005).

A tudományos együttműködések vizsgálata a gyakorlatban gyakran a társszerzőségekre korlátozódik. Ennek oka arra vezethető vissza, hogy a jelenség legegyszerűbben a társszerzőségeken keresztül ragadható meg, természetesen tudva azt, hogy a kettő valójában nem egyezik meg és társszerzőség nélkül is léteznek együttműködések (Schubert és Braun, 1990; Braun et al., 1992; Luukkonen et al., 1993). Ugyanakkor szem előtt kell tartani, hogy azok az adatbázisok, melyek lehetővé teszik az óriási elemszámú mintán végzett vizsgálatokat (pl. Web of Science, Scopus), az együttműködések elemzésének szempontjából csupán a társszerzőségekre vonatkozó adatokat tartalmazzák. Tehát doktori kutatásomban én is a társszerzőségekre koncentrálok: a távolságra vonatkozó vizsgálataimat a szerzői szinten, a kapcsolatok megva-

lósulásának szintjén végzem, mely a legelemibb szint, viszont gyakran csak nehézkesen, nagyfokú adattisztítást követően elemezhető.

Egy-egy kutató munkája során önálló publikációi mellett más és más kutatókkal működik együtt és publikál. Ezek a szerzői körök így egymással átfedésbe kerülnek, tagjaik egy összefüggő hálózatot alkotnak. Ebből következik, hogy a tudományos együttműködések hálózattudományi eszközökkel is vizsgálhatók: az így kialakult csoportosulások sok tekintetben hasonlóak más, alaposan leírt hálózatokhoz, mint amilyen például a világháló szerveződése (Barabási, 2016). A hálózattudomány bevonása újszerű elemekkel, másféle nézőponttal gazdagítja a tudományos együttműködések vizsgálatát, ezáltal pedig új válaszokat adhat az ezekkel foglalkozó tudományág, a tudomány-metria kérdéseire.

1.2. Célok

A tudományos együttműködések vizsgálatának több fontos célja van: ezek közül kiemelendő a jövőbeli együttműködések előrejelzése. Ez történhet meglévő bibliometriai adatok elemzésével, a kutatói kapcsolathálózat feltérképezésével, illetve egy-egy téma köré csoportosuló szerzői klaszterek létrehozásával is. Mindhárom esetben kulcsszerepet kap a távolság kérdése, mely a regionális tudomány egyik kiemelt tércategóriája és a hálózatokon is értelmezhető. Értekezésem egyik várható újdonsága egy olyan távolságindikátor elkülönítése, mely bibliometriai, közelebbről hivatkozási adatok alapján olyan kutatók között feltételez lehetséges jövőbeli kapcsolatot, akik még nem publikáltak közösen társszerzős cikket.

Értekezésem célja a különböző távolságtípusok azonosítása az együttműködések résztvevői között, továbbá az egyes típusok jellemzőinek, valamint lehetséges csoportosítási és mérési módjainak feltárása. További cél a távolság mint regionális tudományi tércategória értelmezése a tudományos együttműködési hálózatokon.

Mint arról már az előző fejezetben is szó volt, a távolság fogalma többféleképpen ragadható meg. Tekinthetjük távolságnak a fizikai térben, kilométerben kifejezett távolságot, két kutató tudásbázisának különbségeit, de akár két struktúra szervezeti felépítésének eltéréseit is. Így értekezésem első kutatási kérdése a következő:

1. Milyen távolságtípusok különíthetők el és ezek hogyan értelmezhetők a tudományos együttműködési hálózatokon?

A távolságtípusok elkülönítése után részletesebben két nagy csoporttal, a társadalmi és a kognitív távolsággal fogok foglalkozni. Előbbi a kutatók személyes kapcsolataira fókuszál, utóbbi pedig kutatási témájuk, háttérbeli tudásuk hasonlóságaira. Mivel a bibliometriai kutatások alapegységei a publikációk, a szerzők közötti távolság azonosításához a publikációk vizsgálata során szerzett adatokat a szerzők szintjén szükséges tovább vizsgálni. Ezzel foglalkozik a második kutatási kérdésem:

2. Hogyan terjeszthető ki szerzői szintre a kognitív távolság?

A vonatkozó szakirodalom áttekintése, továbbá saját kutatásaim során felfigyeltem rá, hogy a széles körben kognitív távolságnak nevezett jelenség valójában több komponensből áll. Ez azt jelenti, hogy ha két kutató között egy elemzés nagy hasonlósági értéket, azaz alacsony távolságot mutat ki, az több tényezőnek is betudható. Erről, az általam immár *teljes* kognitív távolságnak nevezett fogalom kettébontásáról szól a 4. fejezet végén ismertetendő konceptuális modell. A disszertáció harmadik kutatási kérdése tehát:

3. Miként értelmezhető a teljes kognitív távolság?

A konceptuális modellt az 5–7. fejezetben egy kiválasztott minta alapján, különböző hálózatt független és hálózattudományi módszerekkel elemzem. Céлом ezzel az, hogy bemutassam, illetve empirikus adatokkal igazoljam a tudományos együttműködések szereplői közötti teljes kognitív távolság két alkotóelemének, a társadalmi és a tiszta kognitív komponensnek a létezését, ezek egymáshoz, illetve a teljes kognitív távolsághoz való viszonyát. A több szempontból végzett vizsgálat lehetővé teszi a jelenség több irányú megközelítését. Értekezésemben így erre a kérdésre keresem a választ:

4. Hogyan viszonyulnak egymáshoz a különböző távolságtípusok?

Vizsgálati mintámat két tudományág, a közgazdaságtan és a természetföldrajz témájában egy öt éves időtartam alatt megjelent publikációkra korlátozom. A témákat úgy választottam ki, hogy egy társadalom- és egy természettudományi terület adatai is rendelkezésre álljanak. Így módom nyílik az azonosított távolságtípusokat két, eltérő felépítésű együttműködési hálózaton megvizsgálni. A kapott eredmények hasonlóságai a konceptuális modell igazolására is szolgálnak. Az ötödik kutatási kérdés:

5. Van-e különbség az egyes vizsgált szakterületek (természetföldrajz, közgazdaságtan) adatai alapján kirajzolódó mintázatokban?

1.3. Az értekezés felépítése

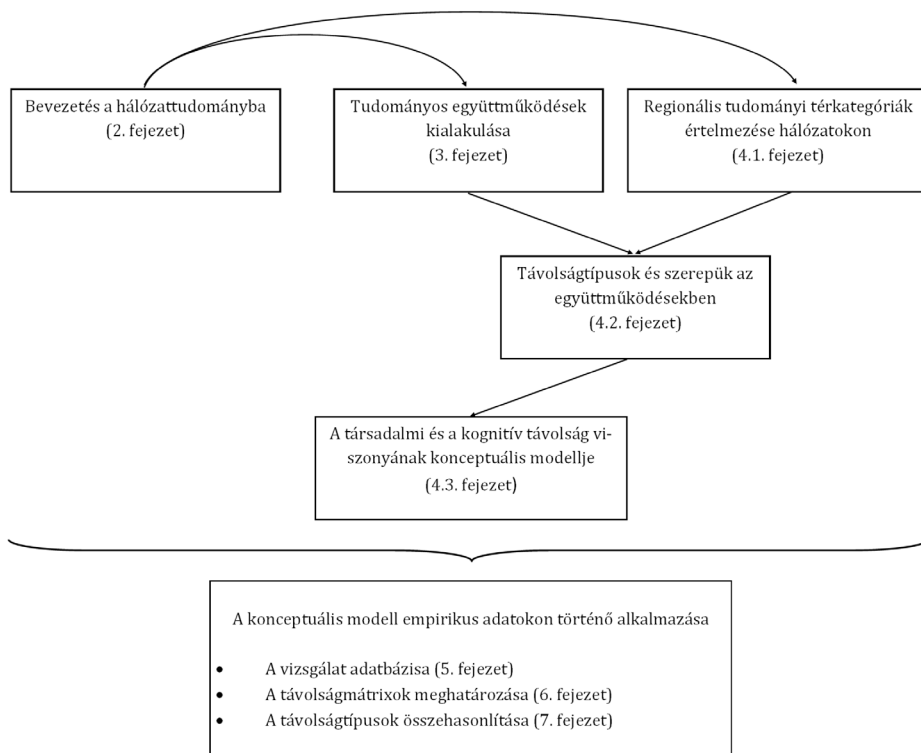
Értekezésemben először elméleti oldalról járom körül a távolság kérdését a tudományos együttműködési hálózatokon. Ehhez először bevezetek bizonyos hálózattudományi alapfogalmakat, melyek ismeretét a későbbi elemzések szempontjából fontosnak tartom. A 2. fejezet tehát röviden bemutatja a hálózattudomány történetének főbb állomásait, jelentősebb elméleti modelljeit. Külön kiemelem, mint a hálózattudomány disszertációm témája szempontjából különösen releváns alkalmazását, a kapcsolatháló-elemzés főbb vizsgálati módszereit, mutatóit.

A 2. fejezetben tehát bevezetésre kerül a hálózat mint egymással kapcsolatban lévő szereplők fogalma. A 3. fejezet az így kialakult együttműködési kapcsolatok kialakulását, jellegzetességeit mutatja be tudományos együttműködésekben. Az „együttműködés” fogalmának tisztázása után egy módszertani szempontból is fontos kérdést, az együttműködés és társszerzőség elkülönítését tárgyalom. Ezek után kitérek a tudományos együttműködések dinamikájára, különös tekintettel létrejöttük okaira, előnyeikre és hátrányaira. A fejezet második felében szakirodalmi példákon keresztül mutatom be az együttműködések vizsgálatának lehetőségeit, melyekben szerepet kapnak hálózattudományi, illetve tudományometriai eszközök is.

A távolság fogalma hozzásegíthet bennünket a tudományos együttműködések jobb megértéséhez, hiszen, mint látni fogjuk, mértéke döntő befolyással van az együttműködések létrejöttére. Ezért a 4. fejezetben ezt a kérdéskört járom körül. A 4.1. fejezetben egyéb térkategóriákkal együtt a távolságot mint regionális tudományi fogalmat ismertetem, illetve a fejezet ezeknek a fogalmaknak a hálózatokon történő értelmezhetőségét is vizsgálja. Ezt követően a 4.2 fejezet a vonatkozó szakirodalom alapján a távolság tudományos együttműködésekben történő értelmezését mutatja be, különböző távolságtípusok, illetve ezek mérhetősége, kutatásokban történő felhasználásuk bemutatásával. Az eddigiek alapján vázolom fel a 4.3 fejezetben azt a modellt, mely egy távolságtípus-csoport, a kognitív távolság komponensekre bontásával pontosabb vizsgálatokat tesz lehetővé a tudományos együttműködési hálózatokon, és megmutatja a kognitív és a társadalmi távolság kapcsolatát.

Az elméleti rész lezárultával a modellt egy konkrét vizsgálatban alkalmazom, két tudományterület, a közgazdaságtan és a természetföldrajz témájában 2010 és 2014 között megjelent publikációk halmazán. Az 5. fejezet a vizsgálatban használt Web of Science adatbázist mutatja be, kiemelve annak előnyeit és korlátait. A rendelkezésre álló adatok számos módon megvilágíthatják a tudományos együttműködések természetét: ebben a fejezetben mutatom be azt, hogy a jelen vizsgálatban milyen bibliometriai adatok alapján határozom

meg és különítem el a kutatásba bevont három távolságtípust. A 6. fejezetben ismertetem, hogy az összehasonlíthatóság érdekében hogyan hozok létre a kapott adatokból egy-egy hasonlósági mátrixot. A vizsgálatokat és eredményeiket a 7. fejezet tartalmazza: a hasonlósági mátrixok adatait különböző hálózattfüggetlen és hálózattudományi módszerekkel elemezve mutatom be a kognitív távolság és komponensei jellemzőit. A két választott tudományág együttműködési gyakorlatának különbözőségeit is ebben a fejezetben tárom fel. A vizsgálatok célja, hogy igazolja a kognitív távolság két komponensének létét. Az utolsó, 8. fejezet az eredmények összefoglalását, annak tágabb kontextusba helyezését tartalmazza, emellett a további kutatás lehetőségeire is kitekintést nyújt. Az értekezés vázlatát az 1. ábra mutatja be.



1. ábra: Az értekezés felépítése

2. Bevezetés a hálózattudományba

Doktori kutatásomban a tudományos együttműködések elemzéséhez a hálózattudomány néhány módszerét is alkalmazom, ezért fontosnak tartom, hogy ebben a fejezetben röviden áttekintsem a hálózattudomány gyökereit és módszertani fejlődésének főbb állomásait. Ezzel egy időben a disszertációban későbbiekben felmerülő, hálózatokhoz köthető fogalmakat és módszereket is bemutatom.

Napjainkban a tudományban egyre nagyobb teret hódít az az elgondolás, hogy a minket körülvevő világ számos jelensége leírható hálózatokkal, illetve hogy bár maguk a jelenségek látszatra igen különbözők, valójában kialakulásukban, felépítésükben, részeik egymáshoz való viszonyában számos hasonlóságot mutatnak (lásd pl. Barabási, 2003, 2016). A hálózattudomány lehetőséget teremt ezen mögöttes szerkezetek feltérképezésére, működési mechanizmusaik megismerésére.

Először a 2.1. fejezetben felvázolom a hálózatok fogalmi alapjait, majd azt, hogy hogyan képezhető le a valóság hálózatokká, illetve ezen struktúrák miként vizsgálhatók. A 2.2. fejezetben egy történeti áttekintést nyújtok. Láthatjuk majd, miként fejlődött ki a hálózattudomány a gráfelméletből, illetve hogy különböző tudományágakban (fizika, matematika, szociológia) tevékenykedő kutatók milyen elméleti modelleket állítottak fel, miként használták a hálózattudomány fogalmait arra, hogy saját tudományágukban új felfedezéseket tegyenek, és ezzel hogyan járultak hozzá a kialakuló hálózattudomány fejlődéséhez.

A hálózattudomány történetének, módszereinek ismertetése alapvető fontosságú a kutatás és a disszertáció szempontjából: a 3.2. fejezetben a hálózattudományi módszerek alkalmazhatóságát láthatjuk a tudományos együttműködések vizsgálatán keresztül, a 7. fejezetben pedig saját kutatásomban is számos olyan fogalmat, módszert alkalmazok, melynek megértéséhez szükséges bevezetést nyújtani ebbe a tudományágba.

2.1. A hálózatok fogalmi alapjai

A hálózatokat nagyon leegyszerűsítve felfoghatjuk bizonyos megfigyelt egységek és az azok között fennálló kapcsolatok összességéeként. Az egységek ábrázolhatók csúcsokként, a kapcsolatok pedig élekként: ebben a megközelítésben minden hálózat váza valójában egy gráf (Mali et al., 2012). E tág definíció is mutatja, hogy az élet számtalan területéről hozhatunk példákat hálózatokra: ilyen az emberi kapcsolatok modellje (társadalmi hálózat), települési, infrastrukturális, közlekedési hálózatok, vízggyűjtő területek leírása

hálózatokkal, vasúthálózat, elektromosvezeték-hálózat (lásd pl. Chorley és Haggett, 1969), gazdasági hálózatok, szervezeti struktúrák, befektetési kapcsolatok hálózatai (lásd pl. Benedek et al., 2007), neurológiai hálózatok és sejtek hálózatai (lásd pl. Böde et al., 2007), társszerzői, citációs hálózatok (lásd pl. Newman, 2004; Leydesdorff, 2007). A felsorolt példák különböző tudományterületek – földrajztudomány, szociológia, közgazdaságtan, biológia – hálózatai, azonban szerkezetükben, topológiájukban, tulajdonságaikban számos hasonlóság fedezhető fel. A hálózattudományi gondolkodás lényege, hogy a valóságot csúcsok és azokat összekötő kapcsolatok egységeként írjuk le, így azt egy gráffal modellezzük. A valóság ily módon történő leegyszerűsítése lehetőséget ad az egyes tudományterületeken elért eredmények más tudományterületek hálózatainak elemzésekor történő felhasználására.

A hálózatok empirikus vizsgálata során az első lépés a hálózat szereplőinek felmérése és ezek egymás közötti kapcsolatainak feltérképezése. A vizsgálni kívánt hálózat elemeinek és kapcsolatainak meghatározásához szükséges előre definiálni, hogy az adott hálózathoz milyen elemeket tekintünk a hálózat csúcsainak, milyen feltételeknek kell teljesülniük, hogy két csúcs között kapcsolat legyen. Például ha telefonhívások alapján kívánunk egy hálózatot meghatározni, akkor egy küszöbértékhez (minimum hívásidőtartam) köthetjük a kapcsolat meglétét (lásd pl. Li et al., 2014).

A hálózatok megadhatók kapcsolati mátrixok, éllisták formájában, de vizuálisan is megjeleníthetők egy gráfként. A hálózattudomány kezdeti időszakában nagyrészt statikus (időbeli dinamika hiánya), irányítatlan (a csúcsok egyenrangúak, nincsenek alá-fölé rendeltségi viszonyok a hálózaton belül) és súlyozatlan (a kapcsolatok erőssége azonos) hálózatokat vizsgáltak: az ilyen hálózatok a mai módszerekkel pontosan elemezhetők, azonban a valóságot nagyobb mértékben egyszerűsítik le. Amennyiben a valóságot pontosabban képezzük le a hálózat megadásakor, és figyelembe vesszük a dinamikát, irányítottságot és a kapcsolatok erősségét, az egymáshoz oly hasonló hálózatok különbözni kezdenek egymástól. Mai módszerekkel, algoritmusokkal azonban a felsorolt három tényezőtől egyszerre maximum kettőt vehetünk figyelembe az empirikus elemzés során ahhoz, hogy akár a legegyszerűbb hálózati paraméterekre vonatkozólag is eredményeket kapjunk (Csermely, 2005).

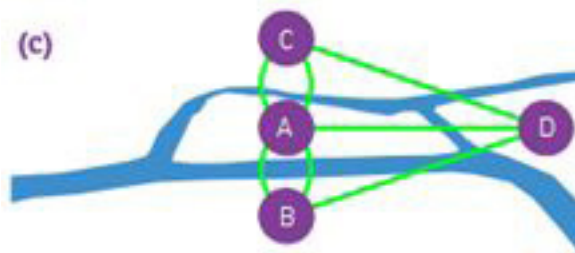
A 20. század végétől került a kutatások fókuszába a hálózatok vizsgálata és a hálózatelemzési módszerek alkalmazása, ezt néhány, hálózatokkal foglalkozó úttörő munka (Erdős és Rényi, 1959; Granovetter, 1973) idézettségének néhány év alatt bekövetkező, ugrásszerű növekedése is jelzi (Barabási, 2016). Napjainkra már önálló hálózattudományról beszélhetünk, ezt jelzi például az új tudományterület intézményesülése is: hálózattudományi PhD képzés indult Budapesten a Közép-európai Egyetemen és az Egyesült Államokban

a bostoni Northeastern Egyetemen. Ezenkívül számos konferencián találunk hálózattudományi szekciót (például a 2017-es MRTT konferencián), ugyanakkor kifejezetten hálózattudománnyal foglalkozó konferenciákat is szerveznek már (egyik legismertebb ezek közül a *NetSci*). Az intézményesülést jelzi a tudományterülethez kötődő saját folyóiratok (pl. a társadalmi hálózatokkal foglalkozó *Social Networks*) megjelenése, másrészt nemzeti tudománypolitikai stratégiák kialakítása is (ld. pl. az Egyesült Államok Országos Kutatási Tanácsának iránymutatása [NRC, 2005]).

2.2. A hálózattudomány fő mérföldkövei

A hálózattudomány közös elméleti gyökerei a matematika (gráfelmélet) és a szociológia (szociometria) területén találhatók, a főbb 20. századi hálózati modellek pedig matematikusoknak és fizikusoknak köszönhetők (pl. Erdős Pál, Rényi Alfréd, Duncan Watts, Steven Strogatz, Barabási Albert-László).

Mi tekinthető a hálózattudomány kezdetének? A hálózatos gondolkodásnak, illetve a hálózatok leírási módjának a gráfelmélet az alapja. A gráfelmélet kialakulása Leonhard Eulerhez és a königsbergi hidak általa levezetett problémájához (2. ábra) köthető. Euler a szigeteket és a partokat csúcsokkal, a hidakat pedig élekkel helyettesítette, így gyakorlatilag egy gráfot írt le, és bebizonyította, hogy ha egy gráfban kettőnél több csúcs rendelkezik páratlan számú éllel, akkor nincsen olyan út a gráfban, ami minden csúcsot és élt csak egyszer érint. Ezzel Euler megoldását tekintik a matematikában az első gráfelméleti megoldásnak (Barabási, 2003, 2016).



2. ábra: A königsbergi hidak gráfként való leképezése

Forrás: Barabási, 2016, 59. o.

A szociológusok szintén régóta, a XIX. század óta foglalkoznak a témával (Newman, 2003). A szociológián belül kiinduló alapnak a kisebb csoportokon belüli kapcsolatokkal foglalkozó tudományterületet, a szociometriát tekintik, valamint a Jacob Levy Moreno nevéhez kötődő, az 1930-as években megalkotott szociogramot (Moreno, 1934). A szociogram a vizsgált személyeket csúcsokkal, a köztük lévő kapcsolatokat pedig vonalakkal ábrázolja. A számí-

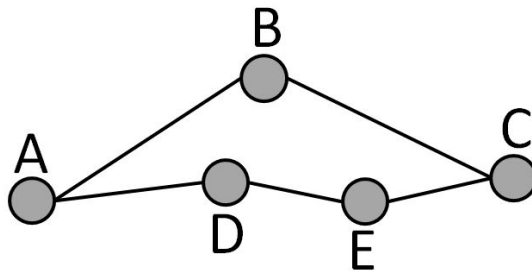
tógépek elterjedése előtt a vizsgálati módszert, a technikai nehézségek miatt, csak kis elemszámú csoportokon lehetett alkalmazni (Tóth, 2009).

A hálózatelemélet következő mérföldkövének az Erdős Pál és Rényi Alfréd magyar matematikusok által létrehozott véletlengráf-modell tekinthető (Erdős és Rényi, 1959). Az Erdős–Rényi modellben adott N db csúcs, és bármely két csúcs között P valószínűséggel szerepel él. A hálózat növekedése során a megjelenő új csúcs P valószínűséggel kapcsolódik egy már meglévő csúcshoz. Az így kialakuló hálózat rendelkezik két alapvető hálózati tulajdonsággal, nevezetesen hogy megtalálható benne az óriáskomponens, valamint érvényes rá a kisvilág-tulajdonság. Ahogy a hálózatban nő az élek száma, egy adott pillanatban az átlagos fokszám eléri, majd meghaladja az egyet ($\langle k \rangle \geq 1$). Ekkor fázisátmenet-szerűen megjelenik az óriáskomponens (Bollobás, 1998; Barabási, 2016). A hálózat óriáskomponense a hálózaton belül a legnagyobb összefüggő alhálózatot jelöli, mely az egész hálózat csúcsainak többségét magába foglalja (Csermely, 2005). Egy hálózaton belül a komponens egy olyan alhálózat, melyben bármely két csúcs között létezik út, azaz összeköttetés.

Annak ellenére, hogy az Erdős–Rényi-modellben létrejövő hálózatot csupán egyetlen valószínűségi paraméter határozza meg, mégis rendelkezik a valós világot leíró hálózatok többségében is gyakran megjelenő kisvilág-tulajdonsággal. A kisvilág-tulajdonság a következőképpen adható meg általános formában:

$$l \sim \log(N)$$

ahol N a csúcsok száma, l az átlagos úthossz. A hálózaton belül a távolság tetszőleges két csúcs között az azokat összekötő úthosszal adható meg: ez az egyik csúcstól a másikig folytonosan egymást követő csúcsokat összekötő kapcsolatok száma. Ezt a kapcsolatszámot lépésnek is nevezik (Barabási, 2016). A 3. ábrán A és C csúcs között két út létezik, az egyik 2, a másik 3 lépés hosszú.



3. ábra: A és C csúcs közötti út

Láthatjuk, hogy egy hálózatban két csúc között több út is található, ilyenkor a két csúc közötti távolságot a köztük lévő utak közül a legrövidebb út hossza jelöli: ez a legrövidebb vagy geodetikus úthossz, melyet általában d -vel jelölnek, azaz jelen esetben A és C csúc között $d = 2$. Fontos megjegyezni, hogy két pont között több egyforma hosszúságú legrövidebb út is létezhet, mely nem tartalmaz hurkot és nem metszi önmagát. A távolsághoz kapcsolódó további alapfogalom a hálózat átmérője (d_{max}), mely a hálózat minden pontpárja között meghatározott legrövidebb utak közül a leghosszabb. A 3. ábrán tehát $d_{max} = 3$. A hálózat jellemzéséhez leggyakrabban az átlagos úthosszt (l) használják, ez a hálózat tetszőleges csúcspárjai között meghatározott legrövidebb utak átlaga (Barabási, 2016, Barabási és Oltvai, 2004). Ez a 3. ábrán $l = 1,5$.

A kisvilág-tulajdonság lényege, hogy nagy hálózatokban is bármely két csúc között a hálózat méretéhez képest is viszonylag alacsony a d értéke. Erre érzett rá Karinth Frigyes is 1929-es *Láncszemek* című novellájában, melyben – még bőven a hálózattudomány megszületése előtt – azt fejtegeti humoros módon, hogy a Föld bármely két lakója között maximum hat lépés hosszúságú ismeretségi lánc található. Milgram (1967) igazolta először kísérletileg ezt a jelenséget, melyet azóta is sokan vizsgáltak (Barabási, 2016). Közülük Duncan Watts (2003) bizonyította és írta le a kisvilág-tulajdonság általános formáját a fent megadott képlettel. Ehhez kapcsolódó érdekes elmélet a három lépés hatótávolság, melyet Nicholas Christakis és James Fowler (2010) ír le könyvében. Elméletük szerint a hálózatokon belül egy csúc a tőle maximum három lépés távolságra levő csúcsra van még hatással.

A következő mérföldkő Duncan Watts és Steven Strogatz kisvilág-modellje (Watts és Strogatz, 1998). Watts és Strogatz azt tapasztalta, hogy a társadalmi kapcsolathálózatok valóban rendelkeznek a kisvilág-tulajdonsággal, de emellett jellemzőjük, hogy klaszterezettek is. A klaszterezettséggel a hálózatokban előforduló háromszög-csoportosulásokat lehet megvizsgálni. A társadalmi kapcsolathálózatokban például gyakran találkozhatunk azzal a jelenséggel, hogy a barátunk barátja a barátunk: ez a gráfban háromszögek formájában jelenik meg. Megkülönböztethetünk lokális, átlagos és globális klaszterezettséget, utóbbi kettő a hálózat egészére vonatkozik (Barabási, 2016). Mindhárom együttható 0 és 1 közötti értékeket vehet fel. A lokális klaszterezettség (C) egy-egy csúc klaszterezettségi mutatóját adja meg. Jellemzően, ha A csúc kapcsolatban van B és C csúccsal, akkor B és C csúc között is lesz él. A lokális klaszterezettség a jelen példában B és C között található él létezésének valószínűségét mutatja, tehát azt, hogy A csúc szomszédjai mekkora valószínűséggel szomszédjai egymásnak is. Képlettel kifejezve:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

ahol C_i az i csúcs lokális klaszterezettsége, L_i az i csúcs szomszédjai között megvalósuló kapcsolatok száma, k_i pedig i csúcs szomszédjainak a száma. Az átlagos klaszterezettség ($\langle C \rangle$) a lokális klaszterezettségi értékek átlaga. Az értéke azt mutatja meg, hogy egy hálózat véletlenszerűen kiválasztott csúcsának két véletlenszerűen kiválasztott szomszédja között milyen valószínűséggel lesz kapcsolat (Barabási, 2016). Az alacsony értékek a véletlen hálózatokra, míg a magas értékek a magasabb hierarchiával rendelkező hálózatokra jellemzők. (Barabási és Oltvai, 2004; Barabási, 2016).

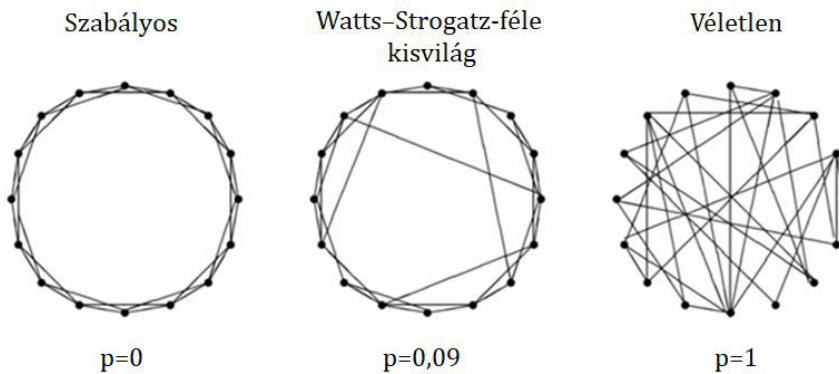
A társadalomtudományokban a 1940-es évekre nyúlik vissza az ún. tranzitív ponthármások arányának vizsgálata, mely mutatót a hálózattudományi szakirodalom globális klaszterezettségnek nevez (Luce és Perry, 1949; Wasserman és Faust, 1994). Ez az érték megmutatja az összefüggő ponthármásokra jutó tranzitív (összefüggő) háromszögek arányát (Barabási, 2016; Newman 2001a).

$$C_{\Delta} = \frac{3(\text{háromszögek száma})}{\text{összefüggő ponthármások száma}}$$

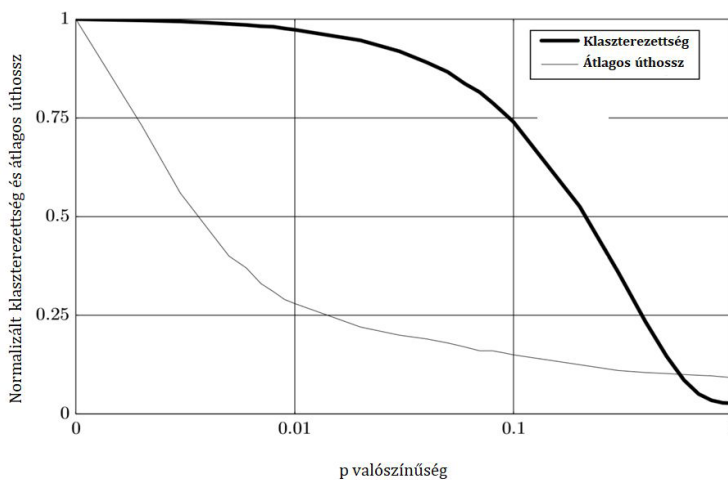
A számlálóban az összefüggő háromszögek szerepelnek: ezt az értéket azért kell 3-mal megszorozni, mert egy ilyen csúcscsoport háromféleképpen írható fel (ABC, BCA, CAB). Mindez természetesen irányítatlan hálózatokra vonatkozik, ahol a kapcsolat két csúcs között vagy létezik, vagy nem, de iránya nincs. A nevezőben az összefüggő ponthármások száma szerepel, mely a zárt ponthármások (háromszögek) és a nyitott ponthármások (olyan ponthármások, amelynek csúcsait két irányítatlan él köti össze) összege. Fontos megjegyezni, hogy a lokális klaszterezettségi értékek egész hálózatra vonatkozó átlaga és a globális klaszterezettség nem ugyanazt az értéket adja.

Watts és Strogatz modellje a szabályos és a véletlen hálózatok között helyezkedik el (4. ábra), a kisvilág- és a klaszterezettség-tulajdonsággal is rendelkezik. Kisvilág-modelljük elnevezése nem összetévesztendő a kisvilág-tulajdonsággal. Modelljük kiindulópontja egy adott $D = 1$ vagy 2 tetszőleges dimenzióban egy szabályos rács, amelyben k szomszédság van (k lehet másod- vagy harmadrendű szomszédság is). Ez a kiinduló hálózat klaszterezett, de kisvilág-tulajdonsággal nem rendelkezik. A kisvilág-tulajdonság megjelenéséhez szükséges a hálózat éleinek módosítása, mely két módon történhet: átdrótozással vagy levágások kialakítása révén. Átdrótozás (rewiring) esetén

4. ábra: Szabályos, Watts–Strogatz-féle kisvilág-, valamint véletlen modell
 Forrás: Cowan, 2004, 9. o. alapján saját szerkesztés



4. ábra: Szabályos, Watts–Strogatz-féle kisvilág-, valamint véletlen modell
 Forrás: Cowan, 2004, 9. o. alapján saját szerkesztés



5. ábra: A Watts–Strogatz-modell tulajdonságai
 Forrás: Cowan, 2004, 10. o. alapján saját szerkesztés

A Watts–Strogatz-modell által leírt hálózat két tulajdonságát ábrázolva (5. ábra) az látható, hogy az átlagos úthossz (l) hamarabb csökken, mint a klaszterezettség (C), ezért lehetséges viszonylag kis $0,001 < [p \text{ átdrótozási valószínűség}] < 0,1$ értékre a klaszterezettség és a kisvilág-tulajdonság együttes jelenléte.

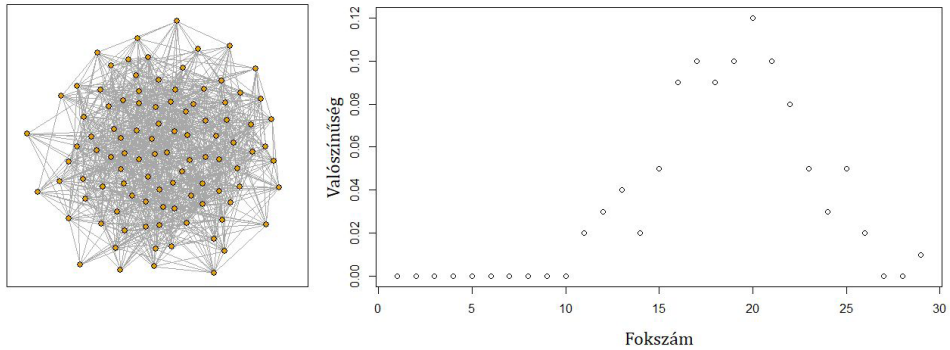
A következő nagy jelentőséggel bíró hálózati alapmodell Barabási Albert-László és Albert Réka által 1999-ben megalkotott Barabási–Albert-modell (Barabási és Albert, 1999). A valódi komplex hálózatok vizsgálata során (pl. az interneten létrejövő új honlapok és a régi honlapok kapcsolatai esetében) arra a következtetésre jutottak, hogy azok fokszámoszlása ún. skálafüggetlen, hatványfüggvény-eloszlást követ. Ezzel szemben a korábbi modellek (véletlengráf-modell, Watts és Strogatz kisvilág-modellje) fokszámoszlása Poisson-eloszlást mutat. Barabási és Albert ennek okát keresve arra a következtetésre jutott, hogy a valós hálózatokban a csúcsok és élek száma folyamatosan változik (pl. az internet esetében új honlapok létrejötte).

A két fő folyamat, mely meghatározza a hálózatok fejlődését, a növekedés és a népszerűségi kapcsolódás. A korábbi modellekben véletlenszerűen kerültek az élek a hálózatokba, és így függetlenek voltak az általuk összekötött csúcsok fokszámaiktól. A népszerűségi vagy preferenciális kapcsolódás azonban figyelembe veszi a hálózat csúcsainak fokszámat. Ez a két mechanizmus skálafüggetlen fokszámoszlással rendelkező hálózatokat eredményez.

A Barabási–Albert-modell a következőképpen adható meg: kiindulásként adott M db kezdőcsúcs, tetszőlegesen összekötve. Minden lépésben hozzáadásra kerül a hálózathoz egy új csúcs (növekedés) E db új éllel, mely a preferenciális csatolás elvét követi, azaz annak a valószínűsége, hogy az új csúcs valamelyik már meglévő csúcshoz fog csatlakozni, arányos az adott csúcs fokszámaival. Az új csúcs tehát a több kapcsolattal rendelkező csúcsot nagyobb valószínűséggel választja szomszédjának. A hálózaton belül ily módon kialakul néhány nagyon sok kapcsolattal rendelkező ún. hub vagy csomópont, de a csúcsok többsége kevés kapcsolattal fog rendelkezni: ez nyilvánul meg a fokszámoszlás hatványfüggvény-eloszlásában (Albert és Barabási, 2002). A Barabási–Albert-modell legfőbb hálózati tulajdonságai a hatványfüggvénnyel leírható fokszámoszlás és a kisvilág-tulajdonság.

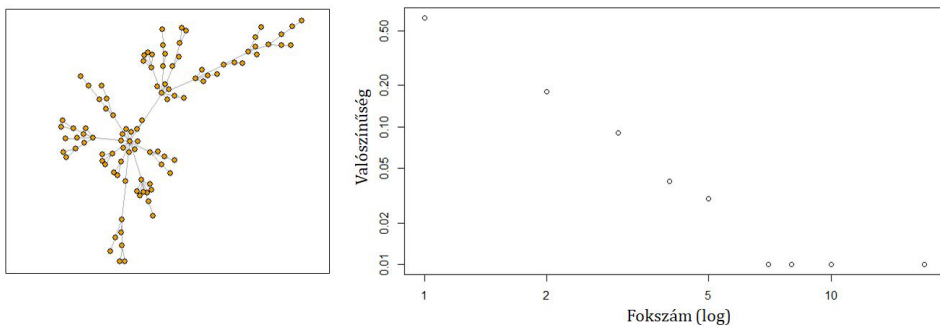
Egy hálózat fokszámoszlása a csúcsok fokszámainak (egy csúcsból kiinduló, illetve beérkező élek számának) gyakorisági eloszlását jelenti. A véletlengráffal leírt hálózat fokszámoszlása Poisson-eloszlással írható le (6. ábra). Az ilyen eloszlással rendelkező hálózatoknál megadható a hálózathoz egy átlagos fokszámoszlás, mely közel áll a maximális fokszámoszláshoz. A valóságban ilyen hálózatokra lehet példa az úthálózat és a vasúthálózat. A skálafüggetlen há-

lózatok (7. ábra) jellemzője, hogy néhány csúcs nagyon sok kapcsolattal rendelkezik, a hálózatok fokszámoszlása így hatványfüggvény-eloszlást követ (Barabási, 2003, Barabási, 2016). Az 6. és 7. ábra a két hálózati alapmodell, az Erdős-Rényi véletlen gráf és a Barabási-Albert-modell skálafüggetlen hálózataát és fokszámoszlását mutatja egyforma méretű, $N=100$ csúcsból álló hálózat esetén.



6. ábra: Erdős-Rényi-féle véletlen gráf és fokszámoszlása

Megjegyzés: $N=100$



7. ábra: Barabási-Albert-moddell létrehozott skálafüggetlen hálózat és fokszámoszlása

Megjegyzés: $N=100$

A valóságban nagyon sok komplex rendszer leírható ilyen hálózatokkal, pl. a repülőterek hálózatai, az internet, a társadalmi kapcsolatháló, a fehérje-hálózatok. A múltban is írtak már le skálafüggetlen eloszlással jellemezhető rendszereket. Zipf (1949) például megállapította, hogy a települések népességszám szerinti eloszlása hatványfüggvény-eloszlást követ. Lotka (1926), majd később de Solla Price (1979) is leírta, hogy a tudományos közlemények eloszlása szintén hatványfüggvénnyel jellemezhető. Ugyanakkor az a tény, hogy ezek a jelenségek hatványfüggvénnyel írhatók le, még nem vonja maga után, hogy skálafüggetlen hálózatokról lenne szó (Nemes Nagy, 2017). Zipf

(1949) megállapítása ugyanis a városok térbeli jellemzőire vonatkozik, míg a hálózatoknál a csúcsok „méretét” kapcsolataik száma határozza meg.

Összehasonlítva ugyanazon fokszámmal rendelkező véletlen és Barabási–Albert-hálózatokat, azt kapjuk eredményül, hogy a Barabási–Albert-modell által leírt hálózat rövidebb átlagos úthosszal rendelkezik, azaz e hálózatokon belül az információáramlás még hatékonyabb (Albert és Barabási, 2002).

Az eddigi modellek és fő hálózati tulajdonságok felismerői és leírói nagyobb részt matematikusok és fizikusok voltak, pedig a hálózatok tudományának „intézményesülése” – például saját folyóiratok (*Social Networks*) megjelenése – egy másik tudományterületen, a szociológiában már régebben megkezdődött. A szociológusok, szociálpszichológusok hálózatokkal foglalkozó munkássága egészen a 19. századig nyúlik vissza. 1940-ben Radcliffe-Brown brit antropológus az elsők között hivatkozott a hálózatra mint a társadalmi struktúrák leképezésének egyik lehetőségére. Az 1980-as évekre már a szociológia egyik fontos területe a társadalmi hálózatelemzés (Varga, 2011).

A szociális hálózatok esetében a hálózat csúcsai az egyes emberek vagy emberek csoportjai, a köztük lévő különböző kapcsolatok pedig az élek. A vizsgálatok középpontjában az emberek, illetve az emberek csoportjai között létrejövő interakciók, kapcsolatok mintázata áll (pl. barátok, cégek, társszerzők, telefonhívások, e-mailek kapcsolathálói). A társadalmi kapcsolatháló-elemzés módszertani előzményeinek Jacob Levy Moreno az 1920-as és 30-as években alkalmazott szociometriai módszerei tekinthetők (Newman, 2003). Az 1950-es évektől kezdtek nagyobb csoportok elemzésére is alkalmas módszereket alkalmazni. Siegfried Frederick Nadel kutatásai fektették le a kapcsolatháló-elemzés (SNA – Social Network Analysis) elméleti alapjait (Tóth, 2009). Az 1960-as években a módszertan fejlődésében hatalmas jelentősége volt a számítógépek megjelenésének, mely lehetővé tette a nagyobb elemszámú hálózatok kezelését, elemzését (Varga, 2011).

A SNA-módszertan segítséget nyújt a társadalmi kapcsolathálóé, belső összefüggéseik megismeréséhez. A kapcsolati háló meghatározását követően különböző mutatókkal jellemezhetjük a hálózat szerkezetét, szereplőinek egymáshoz való viszonyát. Ilyen fontos mutatók és jellemzők pl. a centralitás- és presztízsmutatók, a strukturális egyensúly és tranzitivitás, összetartozó csoportok, modulok, diádok és triádok (Wasserman és Faust, 1994; Scott, 2000; Varga, 2011). Az értekezés későbbiekben felmerülő vizsgálatait szempontjából ezek közül a centralitásmutatók ismertetése szükséges.

A hálózat szerkezetének jellemzése különböző centralitás- (központiság-) mutatókkal is történhet. Több centralitásmutatót is megkülönböztethetünk,

ezek közül a legelterjedtebb a fokszámcentralitás (*Freeman's degree centrality*), a közelségcentralitás (*closeness centrality*), a köztességcentralitás (*betweenness centrality*) és a sajátérték-centralitás (*eigenvector centrality*). Amennyiben a hálózat rendelkezik egy vagy több kiemelten magas centralitással bíró csúccsal, úgy az az egész hálózat hierarchikus felépítésére utal.

A fent felsorolt négy legfontosabb centralitástípus közül az első hármat Freeman (1977, 1979) fejtette ki: eszerint mindhárom típusnak van egy abszolút és egy normalizált értéke. Első helyen áll, egyszerűségénél fogva, a fokszámcentralitás, mely első olvasatban nem más, mint az egyes csúcsok szomszédsági kapcsolatainak összege, azaz a fokszáma. Egy csúcs fokszámcentralitása alapján arra következtethetünk, hogy a magas fokszámú csúcsoknak sok szomszéda van, tehát központi helyet foglalnak el a hálózatban. Normalizálva a fokszámcentralitást, az egyes csúcsokra vonatkoztatva, a fokszám és az elméletileg lehetséges kapcsolatok számának hányadosa:

$$C'_D(p_k) = \frac{\sum_{i=1}^n a(p_i, p_k)}{n - 1}$$

ahol $a(p_i, p_k)$ értéke 1, ha i és k szomszédsági kapcsolatban van, minden más esetben 0.

A következő mutató a köztességcentralitás: ez azzal az értékkel egyenlő, ahány csúcspár közötti legrövidebb út áthalad az adott csúcson. Ez a mutató azt fejezi ki, mekkora közvetítő szerepe van egy adott csúcsnak az információ átvitelében. Az előbbihez hasonlóan ez az érték is függ a hálózat méretétől, így indokolt egy normalizált érték használata, melynek képlete a következő:

$$C'_B(p_k) = \frac{2(\sum_i \sum_j \frac{g_{ikj}}{g_{ij}})}{n^2 - 3n + 2}, i \neq j \neq k$$

ahol g_{ij} az i és j csúcsok közötti legrövidebb utak számát jelöli, g_{ikj} pedig azokat a legrövidebb utakat, melyek áthaladnak k -n.

A harmadik, Freeman által említett centralitástípus a közelségcentralitás: ennek kiindulópontja, hogy azok a csúcsok vannak központi helyzetben, melyek a legkevesebb lépés távolságra vannak a többi csúcstól. Beauchamp (1965) alapján Freeman (1979) a következő képletet adja meg e mutató normalizált értékéhez:

$$C'_c(p_k) = \frac{n-1}{\sum_{i=1}^n d(p_i, p_k)}$$

ahol $d(p_i, p_k)$ az i és k csúcs közötti legrövidebb úthosszt jelenti.

A fokszámcentralitás elvéből indul ki, de más szempontokat is figyelembe vesz a Bonacich (1972, 1987) által kidolgozott sajátérték-centralitás. Eszerint egy csúcs központi szerepét nem csupán fokszáma határozza meg, hanem az is, hogy szomszédai mennyire játszanak központi szerepet: egyfelől a csúcs központiségéhez hozzájárul, ha „erős” szomszédai vannak, de ő maga akkor lesz „erős”, ha szomszédjainak alacsony a fokszáma, tehát függő helyzetben vannak tőle.

Az eddigi centralitásmutatókat jellemzően irányítatlan hálózatokon mérik. Az irányított hálózatok esetében a presztízs fogalma kap nagy szerepet, hiszen ennél a hálózattípusnál az is külön szempont, hogy két csúcs között egy kapcsolat melyik féltől indul ki és ki a „fogadója” (Kürtösi, 2004). Mivel azonban doktori kutatásom témája, a társszerzői hálózatok tipikusan irányítatlanok – tehát két csúcs között a kapcsolat megléte a lényeges – az irányított hálózatok mutatóival a továbbiakban nem foglalkozom.

Az elméletek leírásához kapcsolódó hálózati mérföldkő Mark Granovetter (1973) amerikai szociológus által bevezetett erős és gyenge kötések fogalma. Véleménye szerint valamennyi kapcsolatnak, „kötésnek” megvan a maga szerepe. A kapcsolatok erőssége az idő, az intenzitás, a bizalom és a kötést jellemző kölcsönös „szolgáltatások” függvényében alakul. Egy hálózaton belül erős kötéseknek tekinthetők a közeli, szoros kapcsolatok (Granovetter, 1983). Az ilyen jellegű kapcsolatok gyakran egymáshoz több tekintetben is hasonló szereplők között alakulnak ki. Lehetnek a csúcsok egyedi tulajdonságai hasonlóak, mint például baráti kapcsolatokban egy közös szabadidős tevékenység, hasonló érdeklődési terület, de gyakran az adott csúcsok hálózaton belül elfoglalt helye is hasonló. Gyakran igaz az a megállapítás, hogy a hasonló hasonlót vonz, másképpen fogalmazva a hasonló csoporthoz tartozók között preferált a kapcsolat létrejötte, fenntartása, azaz a társadalmi kapcsolathálózatokban jelen van a homofília (Angelusz és Tardos, 2009). Társadalmi hálózati példát nézve erős kapcsolataink a családunk, barátaink. Az ilyen kapcsolatok esetén az ismerősök között magas az átfedés, magas a klaszterezettség, hiszen gyakran a „barátom barátja a barátom”. Gyenge kapcsolatoknak tekinthetők azok a laza kapcsolatok, amelyek többnyire más-más közegekben mozgó partnerek között jönnek létre. „A hálózat két eleme közötti kapcsolat akkor gyenge, ha a kapcsolat elvétele vagy hozzáadása statisztikusan értékelhető

módon nem befolyásolja a hálózat jellemző tulajdonságának (általában a hálózat egyik csoportjellemző tulajdonságának) átlagát” (Csermely, 2005, 363. o.).

A hálózattudomány fejlődése a számítógépek megjelenése óta nagymértékben felgyorsult. Ebben a gyors fejlődési felfutásban egy heurisztikus környezet jellemezte a hálózattudományt. Egyre több kutató és tudományág mutatott érdeklődést a hálózatok iránt, amint az élet különböző területén leírt hálózatok hasonló tulajdonságokat mutattak. A hálózatkutatás erősen interdiszciplináris volt, és még ma is az maradt. Ugyanakkor ma már (ha például egy konferencia szekcióbeosztását megtekintjük) az egyes tudományterületek a hálózattudományon belül kezdenek egymástól elkülönülni, pl. kapcsolatháló-elemzés, weblapok hálózatai, agykutatás, fehérjehálózatok, hálózati módszerek és a hálózat szerkezetének mind pontosabb leírása, csoportkérésési algoritmusok hálózatokon belül. Mindezek ellenére a hálózattudomány még mindig egy jelentősen interdiszciplináris tudomány, a különböző tudományterületeken tevékenykedő kutatóknak érdemes egymás eredményeit figyelni, mert az egyik területen alkalmazott hálózatvizsgálati módszer kisebb módosításokkal más tudományágakra is átvihető.

3. Tudományos együttműködési hálózatok

A tudományos felfedezések sokszor köthetők több kutató összmunkájához, együttműködéséhez. Ez a tendencia folyamatos erősödést mutat: az 1960-as évektől (pl. Smith, 1958; Clarke, 1964; de Solla Price és Beaver, 1966) egészen napjainkig (pl. Luukkonen et al., 1993; Katz, 1994; Hudson, 1996; Melin és Persson, 1996; Katz és Martin, 1997; Moody, 2004; Sonnenwald 2007) számos kutatás mutatta ki a társ- és sokszerzős cikkek növekvő arányát, főképp a kvantitatív vizsgálatokat tartalmazó munkák körében.

Glänzel és Schubert (2004) megvizsgálták a nem társszerzős publikációk arányát 1980 és 2000 között a Science Citation Indexben (SCI) található folyóiratcikkekre vonatkozóan, és azt találták, hogy míg 1980-ban a cikkek közel 25%-a egyszerszős volt, addig 2000-ben ez az arány 11%-ra mérséklődött. Ezzel párhuzamosan a társszerzős cikkek átlagos szerzőszámát is vizsgálták, mely 2,64-ről 4,16-ra nőtt 1980 és 2000 között. Ez alapján megállapítható, hogy párhuzamosan nem csupán nőtt a társszerzős cikkek aránya, hanem emellett egyre többen is működtek közre egy-egy publikáció megjelenésénél.

3.1. Tudományos együttműködések

3.1.1. Mit értünk tudományos együttműködés alatt?

A tudományos együttműködések vizsgálata előtt azt a kérdést kell tisztáznunk, mit is értünk együttműködés alatt: a kifejezés értelmét ugyanis sokan magától értetődőnek veszik, holott valójában pontos meghatározása nem egyszerű feladat (Laudel, 2001). Az *együttműködés* meghatározása a *Magyar értelmező kéziszótár* szerint: „(az a tény, hogy valaki mással, másokkal) közösen, összhangban, őt v. őket támogatva tevékenykedik, működik” (*Magyar értelmező kéziszótár*, 1972). A *Macmillan English dictionary for advanced learners* pedig a collaborate 'együttműködés' szót a következőképpen határozza meg: „valaminek a létrehozása érdekében együtt végzett közös munkafolyamat” (*the process of working with someone to produce something*) (*Macmillan English dictionary for advanced learners*, 2005). A meghatározásokban közös, hogy több résztvevő közösen tevékenykedik egy közös cél elérése érdekében. Ezek alapján a tudományos együttműködések definiálhatnánk úgy is, hogy az két vagy több személy közös munkája egy kutatási cél elérése érdekében, amely valamilyen új tudományos tudás létrehozását jelenti (Katz és Martin, 1997). Ezt a meghatározást lehet nagyon tágan és szélsőségesen szűken is értelmezni. Az egyik felfogás a kutatói együttműködések extrém tág felfogása mellett érvel, azaz, hogy az egész nemzetközi kutatói közösség egy nagy együttműködés, hiszen minden kutatói együttműködés a tudományos tudáson, egy globális tevékenységen dolgozik (Subramanyam, 1983). Ha ezt egy

globális kutatói hálózatként szeretnénk megjeleníteni, akkor egy teljes hálózatot kell elképzelnünk, amelyben minden csúcson között létezik kapcsolat, mivel a hálózatot alkotó kutatók mindegyike a tudományos tudás létrehozásán tevékenykedik. A másik szélsőséges eset szerint nincs két olyan együttműködő kutató, akiknek teljesen azonosak lennének a céljai, emiatt igazi együttműködés sem létezik (idézi Katz és Martin, 1997). Ezt egy üres, csak csúcsokból álló hálózattal írhatnánk le, melyben nincsenek kapcsolatok a kutatók között. A gyakorlatban a tudományos együttműködések e két szélsőség között valósulnak meg.

Arra a kérdésre, hogy a gyakorlatban, egy konkrét kutatás esetén, ki számít az együttműködésben résztvevőnek, Katz és Martin (1997) tanulmányukban két nézőpont között keresték a megoldást. Az egyik nézet szerint, ha valaki a kutatáshoz hozzátesz valamit a munkájával, már az együttműködők körébe tartozik (tipikus esetei a sokszerzős részecskefizikai tárgyú cikkek); a másik nézőpont azt mondja, hogy az együttműködésben résztvevőknek csak azok tekinthetők, akik az adott kutatás fő részeiben működtek közre. Katz és Martin (1997) öt pontban sorolta fel, hogy véleményük szerint ki számít együttműködőnek.

Együttműködőnek számítanak:

- „azok, akik egy kutatási projekt teljes időtartama alatt vagy jelentős részében közösen dolgoztak, illetve akik gyakori és/vagy jelentős munkával járultak hozzá a projekthez;
- azok, akiknek a neve vagy beosztása szerepel az eredeti kutatási tervben;
- azok, akik a kutatás egy vagy több fő részéért (a kísérlet megtervezése, a mérőeszközök elkészítése, a kísérlet végrehajtása, az adatok elemzése és értelmezése, az eredmények publikálása) felelősek.

Néhány esetben az együttműködők körébe tartoznak még:

- azok, akik egy kulcslépésért felelősek (pl. eredeti kutatási ötlet vagy hipotézis, az elméleti interpretáció);
- az eredeti projekt kitalálója és/vagy finanszírozója, akkor is, ha a továbbiakban a fő hozzájárulása a munkához inkább a projekt menedzselésében (pl. csoportvezetőként), semmint a tulajdonképpeni kutatásban merül ki.

Általánosságban nem tartoznak az együttműködők körébe:

- azok, akik csak alkalmanként vagy viszonylag csekély mértékben járultak hozzá a kutatáshoz;

- azok, akiket nem tekintünk 'valódi' kutatóknak (pl. technikusok, kutatási asszisztensek).” (Katz és Martin, 1997, 7-8.)

Laudel (2001) a tudományos együttműködés következő definíciójából indult ki: „a tudományos együttműködést olyan kutatási tevékenységek rendszereként határozhatjuk meg, melynek szereplői funkcionális módon kapcsolódnak egymáshoz, közös munkájuk során egy olyan kutatási cél elérésén dolgoznak, mely ezen szereplők kutatási témájának vagy érdekeinek felel meg”. Ebből több dolog is következik: az első, hogy az együttműködőknek a kutatási céljai nem feltétlenül egyeznek meg, hiszen például vannak esetek, amikor az egyik közreműködő csak az együttműködési normákhoz való alkalmazkodás miatt segít kollégájának. Másodsorban, Katz és Martin (1997) imént idézett felsorolásával szemben a definíció szerint csak kutatási tevékenység végzése esetén beszélhetünk együttműködésről, így az együttműködők köréből mások (pl. a technikai személyzet) ki vannak zárva. Továbbá egy kutató publikációjára való hivatkozás nem tekinthető együttműködésnek, bár igaz, hogy a kutató munkájának igénybevételével jár. A definícióból kiindulva Laudel (2001) hatféle együttműködést különített el:

- munkamegosztás (a felek egyenrangúságát feltételezi)
- szolgáltatás (nem kreatív hozzájárulás, rutinszerű, ám szakértelmet igénylő feladat elvégzése)
- kutatási felszerelés biztosítása (az együttműködő partner bevezetést nyújt az eszköz használatába, továbbá felel az eszközért)
- know-how átadása (olyan tudás, mely elengedhetetlen a kutatás sikeres kivitelezéséhez, ám leírása nem lehetséges vagy nem szükséges)
- kölcsönös stimuláció (nem a tudás átadása a lényeg, hanem hogy további gondolkodásra, kreatív megoldásokra sarkallja az együttműködőket)
- bizalmi értékelés (a publikáció átnézése, „előlektorálás”).

3.1.2. Tudományos együttműködés és társszerzőség viszonya

Katz és Martin (1997) felhívták a figyelmet arra, hogy a tudományos együttműködés nem egyenlő a társszerzőséggel. Definíciójuk alapján az együttműködők körét az együttműködés folyamatára koncentrálva határozták meg és nem az együttműködés során létrejövő tudományos publikáció létrehozóinak körét írták le. Így a meghatározás egyaránt érvényes egy társszerzős cikkre, egy kutatási projektre, egy közös szabadalomra stb.

A meghatározásból következik az is, hogy nem minden közös kutatási munkafolyamat zárul tudományos publikációval. Ilyen eset például, amikor a kutatási projekt résztvevői külön-külön, a saját tudományterületükön, a saját interpretációjukban publikálják a közös eredményeket. Ugyanakkor arra is találunk példát, hogy egy társszerzős publikáció nem valódi együttműködés eredménye, mert például olyan kutatók is fel vannak tüntetve társszerzőként, akik neve csak presztízsokokból szerepel a szerzők között (például a laboratórium vezetője elvárja, hogy minden, az intézményből származó publikáció társszerzőjeként tüntessék fel), de az is lehetséges, hogy tudományos „csalást” kíván a szerző elfedni fiktív társszerzők bevonásával (részletes esetleírásokhoz lásd Lafollette, 1992).

Ily módon a tudományos együttműködés és a társszerzőség valójában nem azonos, bár a szakirodalomban gyakran a tudományos együttműködések vizsgálata alatt ténylegesen társszerzős publikációk vizsgálatát értik. Szellemes párhuzamot találunk Katz és Martin (1997) cikkében: a tudományos együttműködés közös publikációban való mérése ahhoz hasonlít, mintha a házasságok számát a születési anyakönyvi kivonatokban megjelölt szülőkkel becsülnénk meg. Számos házasság rejtve maradna, viszont sok nem-házasság bekerülne a mintába.

Azok a közreműködők, akik neve megemlízésre kerül publikációkban, ám nem szerepelnek társszerzőként, a köszönetnyilvánításban, esetleg lábjegyzetekben megjelenő alszerzők (*sub-authors*) (Patel, 1973). Az alszerzők különbözőképpen járulhatnak hozzá a publikáció születéséhez. Salager-Meyer és munkatársai (2006) egy hetes felosztást dolgoztak ki, ezek a vizsgált mintán csökkenő gyakorisági sorrendben: anyagi támogatás; eszközbeni, technikai támogatás; konceptuális, akadémiai támogatás; morális támogatás; szerkesztői / nyelvi lektori támogatás; egyéb; irodai támogatás.

Mindezzel együtt is csaknem lehetetlen az egy projekt során megvalósuló tudományos együttműködők teljes körű feltárása. A rejtve maradó együttműködésekhez lásd pl. Laudel (2001): az előző alfejezetben ismertetett együttműködés-típusok vizsgálata alapján arra a következtetésre jutott, hogy a kutatói együttműködések közel felének nincs írásos nyoma a publikációkban.

Ez az oka annak a ténynek, hogy a kutatói együttműködések vizsgálata leggyakrabban társszerzőségek vizsgálatán keresztül valósul meg. A tudományos együttműködések ugyanis legegyszerűbben a társszerzős cikkeken keresztül foghatók meg és mérhetők (Schubert és Braun, 1990; Braun et al., 1992; Luukkonen et al., 1993). A tudományos együttműködések és a társszerzőségek között pozitív korrelációs kapcsolat áll fenn, ahogy arra például Glänzel és Schubert (2004) is rámutat. Így a tudományos együttműködések

vizsgálatával foglalkozó publikációk nagyrészt társszerzős művek elemzésére szorítkoznak, ennek háttérében, az együttműködések mérhetőségén túl, a tudománymetria azon módszertani sajátossága áll, hogy elsősorban bibliometriai adatokat elemez kvantitatív módszerekkel.

A társszerzős cikkek mellett ugyanígy van lehetőség a szabadalmak, vagyis a patentek elemzésére is, mivel a szabadalmak bejegyzésekor azok feltalálói megadásra kerülnek. A patentadatokat a bibliometriai és/vagy citációs adatbázisokhoz hasonló adatbázisok gyűjtik és foglalják magukba, ilyen például a European Patent Office (EPO) és a Patentscope (WIPO) adatbázisa.

3.1.3. Miért jönnek létre tudományos együttműködések?

A tudományos kutatás struktúrája a II. világháborút követően megváltozott, és az ún. „kis tudományt” (*little science*) felváltotta a „nagy tudomány” (*big science*) (de Solla Price, 1979), melyre a kutatók és a publikációk megnövekedett száma, a nagyszabású, országhatárokon átívelő projektek életre hívása, a tudomány nemzetközivé válása jellemző. Ez a megváltozott kutatási környezet hozzájárult az együttműködésekben résztvevők számának a növekedéséhez, egyre gyakoribbá váltak a többszerzős publikációk, a csapatmunkák, valamint egyes területeken (pl. magfizika) a sokszerzős óriás együttműködések (Beaver, 2001). Ziman (1994) egyenesen azt állította, hogy napjainkra a tudományos tudás létrehozásának alapegysége már nem az egyén, hanem a kutatócsoport. Ezekre a csoportokra jellemző, hogy egy csoportvezető (*principal investigator*) felügyelete alatt friss PhD-sok, doktoranduszok és egyetemisták dolgoznak a projekteken (Mali et al., 2012).

A tudományos együttműködések kutatóknak számot kell adniuk ezeknek a – szemünk előtt is folyamatosan zajló – változásoknak az okairól, megjelenési formáiról, valamint egyéb általánosabb jellegzetességeiről. A következőkben az együttműködések ezen jellemzőit veszem górcső alá, a vonatkozó szakirodalom alapján.

A tudományos együttműködések létrejöttének okai

Az együttműködések létrejöttét befolyásoló tényezőkről számos helyen olvashatunk. Katz és Martin (1997) szakirodalmi összefoglalójukban tíz faktorról beszéltek, majd saját kutatásuk alapján ezt hatra szűkítették; Melin (2000) egy kérdőíves vizsgálat során a kutató szükségleteiből indult ki, és három fő faktorcsoportot különített el; Beaver (2001) interjú vizsgálat alapján 18 okot sorolt fel az együttműködésre; Sonnenwald (2007) összefoglaló jellegű munkájában ötféle faktorról beszélt. Ezen csoportosítások alapján én négy fő csoportra osztottam az együttműködések okait, az alábbiakban ezt ismertetem.

Első helyen említendő a *tudományos tényezők* szerepe: ezek azon faktorok, melyek az adott tudományágból, a kutatás elméleti háttéréből fakadnak. Katz és Martin (1997) modelljükben két kulcstényezőt emeltek ki, mint a tudományos együttműködések fontos motorjait: ezek a specializálódás igénye, valamint az interdiszciplináris tudományágak egyre növekvő szerepe. Az együttműködés segíti a komplexebb problémák megoldását, azon jelenségek vizsgálatát, melyeknél egyetlen kutató tudása nem elegendő (Beaver, 2001; Sonnenwald, 2007). Ennek alpontjaként említendő az együttműködés szerepe a *tanítás-tanulás* folyamatában: Melin (2001) a kutató tudásalapú szükségletéről beszélt az együttműködés kapcsán, de Beaver (2001) és Sonnenwald (2007) is amellett foglalt állást, hogy mindenkinek, de különösen fiatal kutatóknak, egyetemi hallgatóknak kiemelkedően fontos az együttműködésekben való részvétel, hiszen így sajátíthatják el az adott tudományág hallgatólagos tudását (*tacit knowledge*).

Az együttműködések létrejöttét nagyban befolyásolják a *rendelkezésre álló erőforrások*: ezek megnyilvánulhatnak eszközök, berendezések, de pénzbeli juttatások formájában is. Láttuk, hogy a „nagy tudomány” fontos jellemzői a nemzetközi projektek: ezek során például egy kutatóintézet munkatársai használhatják a másik kutatóintézet technológiáját, felszereléseit, de arra is találunk példát, hogy nemzetközi finanszírozással valósulnak meg olyan felszerelések, melyeket a nemzeti kutatóintézetek nem engedhetnének meg maguknak (lásd pl. a CERN laboratóriumait) (de Solla Price, 1979). Fontos még megemlíteni a piaci szereplők részvételét az új tudás létrehozásában: egyetemekkel, kutatóintézetekkel együttműködve ezek olyan erőforrásokhoz, pályázati lehetőségekhez juthatnak, melyektől máskülönben el lennének zárva (Autio et al., 1996). Ehhez kapcsolódóan fontos még kiemelni az ún. „hármasspirál” (*triple helix*) modelljét: ez a tudomány–gazdaság–kormányzat hármasságának összefonódását jelenti, mely szereplők intézményei folyamatosan kölcsönösen együttműködnek egymással és pozitív visszacsatolásokon keresztül elősegítik az innovációk megjelenését: ily módon a „hármasspirál” megvalósulása a tudásalapú gazdaság indikátora lehet (Leydesdorff és Etzkowitz, 1998; Lengyel, 2005).

Részben az előző ponthoz kapcsolódik, mégis érdemes külön említeni a *tudománypolitikai tényezőket*: az egyes országok közötti pozitív vagy negatív kapcsolatok befolyásolhatják a nemzetközi tudományos együttműködések létrejöttét, illetve a nemzeti, nemzetközi szervezetek tudománypolitikai eszközökkel befolyásolhatják az együttműködést (lásd pl. Sonnenwald, 2007 összefoglalását). Ilyen tudománypolitikai eszköznek tekinthetők az európai uniós pályázatok (pl. H2020, korábbi FP7, COST), vagy például a hazai tudománypolitikában az NKFIH hazai kutatási projektjei (OTKA), illetve a kiemelt kutatási projektek, mint a Lendület program. Ezek a programok mind elsősor-

ban alapkutatót támogatnak, és a kutatói közösségek egy-egy tématerületen belül szabadon maguk dönthetik el, hogy milyen konkrét téma kutatásával pályáznak. Igaz, az Európai Unió H2020, vagy a korábbi FP7-es projektek esetében előre kiírt tématerületekre kell pályázni, de ezek tág kategóriák és a konkrét kutatási projekt témáját már a pályázói kör dönti el. Melin (2000) interjú és kérdőíves vizsgálatokon alapuló tanulmányában kiemelte, hogy maguk a kutatók sem nézik jó szemmel, ha érezhetően valamilyen irányba akarják terelni őket, vonatkozzon ez akár a kutatás témájára, akár a kutatói együttműködésben résztvevők körére.

Az európai uniós kutatási projekteknél a kutatói konzorcium széles földrajzi lefedettsége szinte elvárt a pályázatok benyújtásakor. A pályázatot eleve egy nemzetközi kutatói konzorciumnak kell benyújtania: előny, ha ennek nyugat-európai partnerországok mellett kelet-közép-európai országok is tagjai. Ez szintén tudománypolitikai ösztönzőnek tekinthető, mely elősegíti a nemzetek közötti kutatói együttműködések elmélyülését. Az Európai Unió, felismerve a nemzetközi kutatói együttműködések fontosságát, olyan programot is létrehozott (COST), amely a nemzetközi tudományos együttműködések létrejöttét hivatott támogatni: ez a program elsősorban egy adott kutatási téma iránt érdeklődő minél szélesebb kutatói kör bevonását célozza pl. konferenciák és workshopok, tanulmányutak, fiatal kutatók nemzetközi tapasztalat-szerzésének a támogatásával (www.cost.eu).

Érdekes kiegészítés az előzőekben látott tényezőkhöz a szerzők, publikációk és támogatások körforgása (Börner et al., 2004). Azok a szerzői körök, akik sok és jó publikációt írnak, nagyobb eséllyel nyernek támogatást a további munkájukhoz. A többlettámogatás elősegíti a magasabb minőségű kutatók, diákok, posztdoktorok bevonását a kutatócsoport tagjai közé, és így több és szintén jó minőségű munkák publikálását.

Egyéni szinten az együttműködések létrejöttében kiemelkedő bizonyos *szociális faktorok* szerepe. Katz és Martin (1997) a tudományt „társas intézményként” írta le, melyben az egyén előmenetele, láthatósága nagymértékben függ a többi kutatóval történő interakcióktól. Ez a felfogás összhangban áll Subramanyam (1983) már említett koncepciójával, melyben az egész tudományos közösség felfogható egyetlen hatalmas hálózatként. Többen (ld. pl. Beaver, 2001; Sonnenwald, 2007) arra fektették a hangsúlyt, hogy új együttműködések létrejötténél az adott kutató a saját kapcsolati hálójából indul ki, legyen szó akár projektek, akár új partnerek felkutatásáról. (A tudományos együttműködések hálózattudományi vizsgálatához lásd bővebben a 3.2. fejezetet.) Az intézmények közötti együttműködések kialakításának fontos színterei a nemzetközi konferenciák, workshopok, ahol lehetőség adódik mind a formális, mind az informális eszmecserére; épp ezért tudománypolitikai, tu-

dományfinanszírozási szempontból is igazán fontos olyan előnyös feltételek megteremtése, melyek segítik a kutatók külföldi tapasztalatszerzését (Melin, 2000). A szociális faktorok alcsoportjaként említhetők bizonyos *személyiségbeli tényezők* is: az együttműködést elősegíti a téma iránti közös érdeklődés, a kutatás élvezete (Beaver, 2001), illetve a kölcsönös szimpátia (ld. pl. Melin, 2000).

A tudományos együttműködések előnyei

Arisztotelésztől kezdve sokan hangoztatták: „az egész több, mint a részek összessége”. A tudományos együttműködésekre is fokozottan igaz ez a tétel, hiszen a közös munka során a különböző tudások összeadásából valami új jön létre, az ötletek cseréje kreatív ötleteket hív életre, ráadásul a felek is új tudás birtokába jutnak egymástól.

Elsőként kell megemlítenünk az együttműködés *hatékonyságával* kapcsolatos előnyöket. Az együttműködés során az egyes kutatók hatékonyabban használják ki szakértelmüket, hiszen ideális esetben mindenkinek csak a saját szakterületével kell foglalkoznia, legyen az adatfeldolgozás vagy az eredmények disszeminációja (Katz és Martin, 1997). *Kreatív* előnye az együttműködésnek, hogy a különböző nézőpontok ütköztetése, az együtt gondolkodás segíti az újdonság megszületését (Beaver, 2001).

Ha több kutató együttműködik, a közös munka során minden résztvevő olyan képességek, tudás birtokába jut, melyekkel eddig nem rendelkezett. Ez talán a másik legfontosabb előnye az együttműködésnek: segíti a *tudás áramlását* (Katz és Martin, 1997).

Az együttműködés *„logisztikai”* szempontból is egyértelműen számos előnnyel jár, ezek közül megemlíthetjük a gyorsaságot (szét lehet osztani a munkát több ember között), a széleskörűséget (pl. egyetemi hallgatók bevonásával nagyobb mintát lehet vizsgálni), illetve a „több szem többet lát” elvén a hibák gyorsabb, hatékonyabb kiszűrését (Beaver, 2001).

A *láthatóság, kapcsolatépítés* is fontos előnyként fogalmazódik meg. A többszerzős cikkek idézettsége nagyobb (Narin et al., 1991; Goldfinch et al., 2003; Frenken et al., 2005), továbbá nemzetközi együttműködés esetén minden szerző a saját országában tudja disszeminálni az eredményeket, növelve ezzel az ismertséget (Nomaler et al., 2013). Amint már láttuk, az együttműködés a kapcsolati hálóval kezdődik; ugyanakkor a közös munka során a háló bővülése, a kapcsolatok kiterjesztése, elmélyítése is fontos hozadékként jelenik meg (Katz és Martin, 1997; Melin, 2000).

A tudományos együttműködések hátrányai

A fent ismertetett előnyök mellett olyan tényezők is felmerülhetnek a tudományos együttműködések során, melyek inkább hátráltatják az eredményes munkát. Katz és Martin (1997) ezeket az együttműködés költségeinek nevezték. Ők a következő hátrányokat sorolták fel:

- az egyik legfontosabb hátrány az együttműködés *anyag*i vonzata. Intézetek, nemzetek közötti együttműködés esetén gyakori, hogy a résztvevők elutaznak egymáshoz, anyagokat, felszerelést küldenek egymásnak. Ez mind-mind pénzbeli költséggel jár, melyre nincs mindig megfelelő fedezet a kutatás finanszírozásában;
- amiatt, hogy az együttműködők sok esetben nem egy helyen dolgoznak, megnövekedett *időre* van szükség, az utazás, a csomagok küldése miatt. A különböző nézőpontok egyeztetése, a felmerülő ellentétek kezelése is több időt igényel;
- a megnövekvő *adminisztráció*, legyen az utazások miatti ügyintézés, pályázati menedzsment, vagy a kutatási projekt ügyintézése.

Az együttműködés költségeként említhetjük még Beaver (2001) megállapítását, mely szerint egy többszerzős cikknél csökken az egyén láthatósága, így az egyén szempontjából nem feltétlenül előnyös: elég arra gondolni, hogy tudományáganként változó annak gyakorlata, ki kerül az első helyre egy többszerzős cikk szerzői közül – tehát, hogy kinek a neve szerepel biztosan a hivatkozásokban – illetve ki kerül egyáltalán a szerzők közé (Dance, 2012).

Az együttműködések a különböző tudományágakban

A tudományos együttműködések kialakulása az előzőekben felvázolt, általános tényezőkhöz kívül attól is függ, hogy egy adott kutató milyen tudományterületen tevékenykedik. A disszertáció empirikus részében mi is látni fogjuk részletesen, hogy különbségek találhatók az együttműködések jellemzői terén különböző tudományágak között, ezért itt csak röviden tárgyalom a témát. Smith (1958) számolt be először arról a jelenségről, hogy az elméleti publikációk jellemzően kevesebb szerzővel rendelkeznek, mint a kísérleti témájúak (Meadows és O'Connor, 1971; de Solla Price, 1979; Gordon, 1980), a természettudományokban pedig gyakoribbak a társszerzőségekre, mint a társadalomtudományokban (Mali et al., 2012). Hasonló jelenséget írt le többek között Newman (2001a) is, aki több adatbázis vizsgálata során az elméleti részecskefizika és a számítógép-tudomány témájában született publikációk 1,99, illetve 2,22-es átlagos szerzőszámát állította szembe például az orvosi-biológiai publikációk 3,75-ös átlagos szerzőszámával. Ezzel együtt nem sza-

bad megfelelkezni arról a már említett tényről, hogy tudományágtól függetlenül az egyes publikációk átlagos szerzőszáma különböző mértékben bár, de növekvő tendenciát mutat (lásd pl. Glänzel, 2002; Mali et al., 2012).

3.2. A kutatói együttműködések empirikus vizsgálata

A kutatói együttműködések vizsgálata a gyakorlatban leggyakrabban a társszerzőségek elemzésére korlátozódik. Ez az együttműködések mérhetőségére vezethető vissza, ahogy azt a 3.1.2. fejezetben láthattuk. A kutatók többsége, amikor tudományos együttműködések elemzését csupán a társszerzőséget érti alatta, illetve csak azokat vonja be a vizsgálatába (lásd pl. Schubert és Braun, 1990; Newman, 2001a-c, 2004; Ponds et al., 2007; Acosta et al., 2011).

Maguk az együttműködések elsődleges közreműködői a kutatók, azonban az együttműködések vizsgálata több szinten is történhet a kutatók affiliációi (kutatóhelyei) alapján (Melin és Persson, 1996). Hou és munkatársai (2008) tanulmányukban három szintet különítettek el ebből a szempontból: a mikroszinten a szereplők maguk a szerzők, a mezoszinten a kutatócsoportok, az intézmények, míg makroszinten a nemzetek, országok az együttműködések szereplői. Doktori kutatásomban én a mikroszintet elemzem, mivel az együttműködések a társszerzőségeken keresztül vizsgálom.

Ebben a fejezetben a tudományos együttműködések empirikus vizsgálatával foglalkozó, eltérő kutatási fókuszú tanulmányokból mutatom be az általam leglényegesebbnek vélteket. A tanulmányozott szakirodalmat három nagy csoportba soroltam. Az alkalmazott módszerek szerint elkülönítettem (1) egy tisztán hálózatelemzési módszereket használó, hálózatos szemléletű csoportot, (2) egy vegyes, nem kizárólag hálózatos fókuszú, valamint (3) egy olyan csoportot, mely a kutatói együttműködések vizsgálatát némileg más szemszögből, a *távolság* fogalmából kiindulva végezte.

Az (1) csoportba sorolt tanulmányok (Newman, 2001a-c, 2004; Barabási et al., 2002; Börner et al., 2004) különböző tudományterületek társszerzői hálózatát vizsgálták, bemutatták a vizsgált hálózatok alapstatisztikáit és tulajdonságait (pl. klaszterezettség, átlagos úthossz, centralitásmutatók). Newman tanulmányai egy időszakra vonatkozóan statikusan mutatták be a hálózatokat, míg a másik két vizsgálat az időbeliséget is figyelembe vette dinamikus modellek alkalmazásával.

A (2) csoportba sorolt tanulmányok (Braun et al., 1992; Luukkonen et al., 1993; Melin és Persson, 1996; Heimeriks et al., 2003; Wagner és Leydesdorff, 2005; Hou et al., 2008) a hálózatos módszerek mellett más módszerekkel is

vizsgálták a társszerzői hálózatokat, a kapott eredményeket pedig sokkal inkább a tudománymetria területén interpretálták.

A (3) csoportba tartozó tanulmányok (Katz, 1994; Autant-Bernard et al., 2007; Ponds et al., 2007; Mattsson et al., 2008; Frenken et al., 2009a; Hoekman et al., 2009; Hoekman et al., 2010; Acosta et al., 2011; Nomaler et al., 2013) nem alkalmaztak hálózattudományi módszereket, igaz, ezekben a tanulmányokban a kutatási fókusz az együttműködések vagy a társszerzőségeken volt és nem a társszerzői hálózatokon. Ezek a tanulmányok a kutatók közötti viszonyokat különböző tényezők mentén vizsgálják, és gyakran regresszióanalízissel próbálják megállapítani, hogy vajon melyik tényező az, mely leginkább meghatározza az együttműködések során létrejött publikáció sikerességét. A sikerességet valamilyen, a tudománymetriában használt hatásmutatóval mérték (pl. a hivatkozások száma). A (3) csoportba sorolt tanulmányokat a 4.2 fejezetben veszem górcső alá, az együttműködésben résztvevő szereplők közötti távolsághoz kapcsolódó tényezőkre fókuszálva.

A kutatói együttműködési hálózatok kapcsolathálózatként értelmezhetők, melyekben a csúcsok, vagy ahogy a kapcsolathálózat-elemzés során nevezik, a szereplők a kutatók, a közöttük lévő kapcsolatok vagy élek pedig az együttműködések, illetve a közös publikációk (Newman 2001a). Az (1) és (2) csoportba sorolt tanulmányok tehát a kutatók kapcsolathálózatait vizsgálták a társszerzői hálózatokon keresztül.

3.2.1. Az együttműködések hálózatelemzési módszerekkel vizsgáló tanulmányok

Newman (2001a-c, 2004) több tanulmányában is a társszerzői hálózatokat elemezte mikroszinten, azaz a kutatókra fókuszálva. 2001-es cikksorozatában az orvosbiológia, fizika és számítástechnika, míg 2004-es tanulmányában az orvosbiológia és a fizika mellett a matematika területén megjelent folyóiratcikkek társszerzői hálózatát vizsgálta alapvető hálózati mutatók meghatározásával. Az általános gyakorlattal ellentétben nem a Science Citation Index (SCI) vagy később a Web of Science (WoS) adatbázist (A WoS adatbázist a 5.2 fejezetben mutatom be részletesen), hanem külön-külön egy, az adott tudományterülethez köthető adatbázist használt. Az alapstatisztikák közül megnézte az *egy cikkre jutó szerzőszámot*, az *egy szerzőre jutó cikkszámot*, az *egy együttműködésre jutó szerzőszámot* és az *együttműködők átlagos számát*. E mutatószámok eloszlása közel hatványfüggvény-eloszlást követ. Jellemző a hosszú „farok” megléte, ami például a szerzőkre jutó cikkek számánál azt jelenti, hogy néhány szerzőhöz nagyon sok publikáció köthető, míg sok szerző csak néhány cikket publikált a vizsgált öt éves időablakban. A valós adatok fokszámoszlásának elemzésénél, illetve az együttműködők számának el-

oszlását vizsgálva Newman (2001a, 2001c, 2004) nem talált tiszta hatványfüggvény-eloszlást, így skálafüggetlennek sem nevezte a vizsgált társszerzői hálózatokat, mivel nem teljesült minden ehhez szükséges feltétel (Barabási, 2016). Egy módszertani eljárást, az exponenciális vágást alkalmazva ugyanakkor hatványfüggvény-jellegűnek tekinthetők az eloszlások. Ezt Newman a választott időablak (5 év) rövidségével magyarázta. Ez így tulajdonképpen tipikus esete annak, amikor a valóságot leíró adatokat egy alapmodellel (skálafüggetlen hálózat) hasonlítjuk össze.

A Newman (2001a, 2001c, 2004) által vizsgált területek társszerzői hálózata rendelkezik *óriáskomponenssel*. Az orvosbiológia területén a csúcsok több, mint 90%-a található meg a hálózat legnagyobb összefüggő komponensében, az óriáskomponensben, de a többi területen is magas arányok a jellemzők.

Az egyes tudományterületek között az együttműködők átlagos számát tekintve a gyakorlati tudományok területén magasabb értékek voltak kimutathatók, mint az elméleti területeken (Newman, 2001a). A legnagyobb átlagos szerzőszám a részecskefizika (11,6) területén volt tapasztalható, amit az orvosbiológia területe követett (3,75) (Newman 2001a, 2001c, 2004). A részecskefizika területén akár több ezer szerző is köthető egyes publikációkhoz, emiatt ezt a területet érdemes külön vizsgálni, mivel itt a „szokásos” (2-3 fő) társszerzői létszám sokszorosa a jellemző.

A lokális, átlagos és globális klaszterezettségmutatók közül a két utóbbi a hálózat egészére ad információt. Newman (2001a, 2001c, 2004) ezek közül a globális klaszterezettséget vizsgálta tanulmányaiban: a mutató kidolgozása a társadalomtudomány területén kapcsolathálózat-elemzéssel foglalkozókhoz köthető (Luce és Perry, 1949; Wasserman és Faust, 1994). A szerző viszonylag magas értékeket talált a globális klaszterezettségre vonatkozóan (0,3 és 0,7 között), kivéve az orvosbiológia esetében, ahol jóval alacsonyabb, 0,07 értéket kapott. A viszonylag magas értékeket a szerző azzal magyarázta, hogy gyakoriak a három- vagy többszerzős cikkek, és ebben az esetben egy-egy cikk szerzői a társszerzői hálózatban háromszögeket alkotnak. Ám, mint arra rámutatott, ez csak az esetek kb. 50%-ára adhat magyarázatot, a többi esetben más, szociális okokat kell keresnünk. Ezek a következők lehetnek:

- egy kutató két kollégájával külön-külön együttműködik, később ennek hatására a két kolléga is közös munkába kezd. Ez a hálózati tranzitivitás tipikus esete;
- három kutató hasonló körökben mozog (ugyanazt a folyóiratot olvassák, gyakran találkoznak konferenciákon), és ennek hatására páronként külön-külön együttműködés alakul ki közöttük, bár hárman közös kutatást nem végeznek;

- az előző eset speciális példája, amikor a három kutató ugyanabban az intézetben dolgozik, így merül fel közöttük a páronkénti együttműködés lehetősége.

Newman (2001b) következő cikkében az előző tanulmányát folytatta, a vizsgált társszerzői hálózatokban további szerkezeti paramétereket vizsgált, úgy mint a távolság és a legrövidebb úthossz, valamint a különböző centralitásmutatók. Ezek a mutatók a hálózat szereplői közötti kapcsolatok megismerését segítik, így módon a kutatók közötti kommunikációs és kapcsolati mintázatok megismerését teszik lehetővé. A szereplők közötti legrövidebb utakat vizsgálva megmutatta, hogy a hálózatok rendelkeznek az ún. kisvilág-tulajdonsággal, azaz tetszőlegesen választott két csúcs között viszonylag kis távolságot találunk a hálózat méretéhez (csúcsok száma) képest. Az átlagos úthossz a hálózat méretéhez viszonyítva rövidnek bizonyult, a matematika területe (Newman 2004) és a számítástechnika (Newman 2001b, 2001c) mutatta a legmagasabb értéket; ez egyben az a társszerzői háló, ahol az együttműködés átlagos száma a legalacsonyabb volt.

A preferenciális csatolással létrejött hálózatok fokszámoszlása hatványfüggvény-eloszlást követ (Barabási és Albert, 1999; Newman, 2003), azaz a társszerzői hálózatok esetében néhány szerzőnek sok társszerzője van és úgynevezett csomópontot (hub) alkot, míg a többség néhány kapcsolattal rendelkezik. Ez a hálózati szerkezet a csomópontokban található központi helyzetű kutatók tudományos elméleteinek az elterjedését nagyban segíti. A kapcsolatszám vagy fokszámcentralitás ahogy a hálózatban, úgy a tudományos életben betöltött státuszra is közelítést adhat (Moody, 2004).

A köztességccentralitást vizsgálva Newman (2001b, 2004) azt találta, hogy egy szerző és a többi szerző közötti utak nagy része az adott szerző egy vagy két társszerzőjén halad keresztül. Ezt Newman Strogatz közlése alapján plasztikusan „tölcséreffektusnak” nevezte.

Newman tanulmányainak (2001a-c, 2004) bevallott célja, hogy bemutassa a hálózattudomány módszereinek alkalmazhatóságát bibliometriai adatokon. Munkái kiváló lehetőséget teremtenek a legfontosabb hálózatelemzési fogalmak ismertetésére a társszerzői hálózatokon keresztül. Ezen fogalmak, mutatók mentén megragadhatók bármilyen hálózat alapvető tulajdonságai; ezt dolgozatomban én is bibliometriai adatokon keresztül fogom bemutatni. Így számszerűsíthető módon hasonlíthatók majd össze a 4. fejezetben általam definiált távolságtípusok alapján meghatározott hálózatok.

Barabási és munkatársai (2002) a statikus társszerzői hálózatok vizsgálata helyett a hálózatok fejlődésére helyezték a hangsúlyt. Véleményük szerint a

komplex hálózatok időbeli változásának tanulmányozására az együttműködések kiválóan alkalmasak, két okból kifolyólag: viszonylag pontos adatok állnak rendelkezésre az új kapcsolatok létrejöttéről (a SCI-nek, később a WoS-nak köszönhetően), továbbá a kapcsolatok létrejötte a szerzők (a hálózat csúcsai) akaratától függ – szemben például egy másik jól dokumentált hálózattal, a hollywoodi színészek közös filmekben való szereplésével, ahol az „együttműködés” rendszerint külső, rendezői vagy produceri döntés eredménye. Vizsgálatukban idegtudományi és matematikai tárgyú, 1991 és 1998 között publikált cikkekből kialakult társszerzői hálózatokat vizsgáltak. A következőkben felsorolt jellemzők mindkét tudományág alakulásában kimutathatók voltak, bár az idegtudományban nagyobb arányban:

- a csúcsok (szerzők) és az élek (társszerzős cikkek) száma folyamatos növekvő tendenciát mutatott;
- az új kapcsolatok létrejöttét a preferenciális kapcsolódás szabályozta mind új belépők (csúcsok) esetén, mind már meglévő csúcsok között;
- a fokszámeloszlás hatványfüggvényt követett, a hálózatok skálafüggetlenek voltak;
- a klaszterezettségi együttható a két időpont között csökkent;
- az óriáskomponens relatív nagysága nőtt;
- az átlagos fokszám nőtt;
- az átlagos úthossz az időben előre haladva csökkent.

Ez a szerzők szerint ellentétes a korábbi modelleknél tapasztaltakkal, az okát abban feltételezték, hogy egyrészt a már meglévő csúcsok között új kapcsolatok alakulnak ki, melyek csökkentik a legrövidebb utat, másrészt, hogy nem áll rendelkezésre 1991 előtti adat, ami viszont az adatbázis torzítását okozhatja.

A szerzők a kapott eredményekből egy általános modellt vázoltak fel a hálózatok evolúciójára, és hangsúlyozták, hogy kutatásuk általános, bevezető jellegű. Szintén dinamikus modellt alkalmaztak Börner és munkatársai (2004) egy szimulációs modellben, mely a publikációk és kutatók közötti viszonyt egyszerre vizsgálta a társszerzői és a citációs hálózatokon keresztül. Elméletük szerint a kutatók és a tudományos publikációk „ikerhálózata” kölcsönösen határozza meg egymást: a citációs hálózatokban a hivatkozásokon keresztül a kutatók kötik egymáshoz a publikációkat, míg a társszerzői hálózatokban a közös publikációk kapcsolják össze a kutatókat. Itt kell megjegyezni, hogy a citációs hálózatok a társszerzői hálózatokkal ellentétben nem kapcsolat-

hálózatok, mivel szereplőik az egyes publikációk. A szerzők modelljüket a PNAS folyóirat bibliometriai adataival validálták. A modell a preferenciális kapcsolódáson alapszik: a szerzők meglátása, hogy napjainkban, a tudomány folyamatos specializálódásával, az egyes kutatóknak nincs rálátásuk a társszerzői és citációs hálózatok egészére, a komplex hálózatnak csupán egy helyi szintjét ismerhetik. A modell által meghatározott hálózat nagyjából skálafüggetlen, az attól való eltérést a tudományos közösség tématerület szerinti megosztottsága és az „öregedés” tényezője magyarázhatja a szerzők szerint. Az „öregedés” alatt a szerzők egyrészt azt értik, hogy egy magasan hivatkozott publikációnak nehéz ezt a magas értéket évről évre fenntartani, a társszerzői hálózatból pedig idővel egy-egy szereplő kiesik.

3.2.2. Az együttműködések részben hálózatelemzési módszerekkel vizsgáló tanulmányok

A (2) csoportba sorolt tanulmányok az együttműködések nem kizárólag hálózatelemzési módszerekkel vizsgálták, más tudományágak, elsősorban a tudománymetria eszköztárából is merítettek. A tanulmányok egy része mikroszinten vette górcső alá a társszerzői hálózatokat (Hou et al., 2008), mezoszinten az intézményi együttműködési hálózatokat (Melin és Persson, 1996; Heimeriks et al., 2003) és makroszinten (Braun et al., 1992; Luukkonen et al., 1993; Wagner és Leydesdorff, 2005) az országok közötti kapcsolatokat.

Hou és társai (2008) a *Scientometrics* folyóiratban megjelent publikációk társszerzői hálózatát elemezték többféle nézőpontból. A társszerzői hálózaton a társadalmi kapcsolatháló-elemzés (SNA) módszereivel vizsgálták, ezenkívül a címekben előforduló szavakon az együtt előfordulás (*co-occurrence*) módszerét, valamint klaszteranalízist és gyakoriságelemzést is végeztek. A társszerzői hálózat fő paramétereinek megállapítása után a legalább 3 cikket publikáló szerzők társszerzői hálózatát jelenítették meg irányítatlan, súlyozott formában, és klaszteranalízis segítségével szerzői csoportokat hoztak létre, emellett a szerzők fokszámcentralitását is ábrázolták. A klaszteranalízis eredményeként a kialakult szerzői csoportok cikkeinek címszávaiból az adott klaszter fő témái is elemzésre kerültek. A hálózat szereplőire a fokszámcentralitás mellett a közelség- és köztességcentralitást is megvizsgálták. Mindegyik esetben ugyanaz a hálózati szereplő rendelkezett a legmagasabb centralitásértékkel. A centralitásértékeket a maximumra normálták, ennek a szerzői kibocsátással való összefüggését vizsgálták meg korreláció segítségével, melyből kiderült, hogy a centralitás és a kibocsátás pozitív korrelációs kapcsolatban van egymással.

Melin és Persson (1996) felhívták a figyelmet a bibliográfiai adatokban rejlő lehetőségekre, és társszerzőségeken keresztül vizsgálták az együttműködé-

seket. Tanulmányukban hangsúlyozták, hogy a társszerzőségek vizsgálata tudománypolitikai szempontból is hasznos. Kutatásuk során egy svéd egyetem társszerzői hálózatát elemezték, melyhez elsősorban tudományometriai eszközöket alkalmaztak. Az intézmény társszerzői kapcsolatai révén bemutatták az egyes tudományterületeken más, nem akadémiai szférával megvalósuló együttműködések arányát és a svédországi egyetemek nemzetközi tudományos beágyazottságát. A tanulmányban hálózatok csak érintőlegesen jelentek meg, csupán a vizsgált egyetem publikációinak nemzetközi társszerzői hálózatának ábrázolásával.

Heimeriks és munkatársai (2003) a tudományos együttműködések több platformon történő vizsgálatára tettek egy érdekes kísérletet. Leydesdorff és Etzkowitz (1998) a tudomány–kormányzat–gazdaság kapcsolatát leíró „hármasspirál” modelljéből indultak ki, és arra keresték a választ, hogy a modern tudományos együttműködésekben milyen új színtereken zajlik a kommunikáció, illetve az ezek nyomán kialakult hálózatok milyen kapcsolatban vannak a „hagyományos” társszerzői hálózatokkal. Vizsgálatukban a tudományos–kormányzati–gazdasági intézmények közötti kommunikációt több módon is próbálták azonosítani az elemzéshez. Egyrészt a bibliográfiai csatolás módszerével (lásd 4.2.3. fejezet) kapcsolatot teremtettek az intézmények között, amennyiben azok azonos irodalmat használtak fel a publikációkban, másrészt kutatási projektek résztvevő intézményeinek együttműködéseit azonosították, valamint az előző két módon azonosított intézményi honlapok egymásra hivatkozásait vizsgálták. A számos módszertani nehézség mellett, eredményeik közül kiemelendő, hogy a „második típusú tudástermelésben” (Gibbons et al., 1994) is az akadémiai szféráé a vezető szerep, ugyanakkor az európai uniós pályázatokban egyre növekvő arányban jelentek meg ipari, piaci szereplők is. Az együttműködésekben résztvevők honlapjait vizsgálva arra a következtetésre jutottak, hogy az internet, mint az eredmények disszeminációjának helyszíne, inkább a lokális célközönségnek szól.

Luukkonen és munkatársai (1993) tanulmányukban elsősorban az országok közötti tudományos együttműködésének vizsgálata során felmerülő módszertani problémák megoldására tettek kísérletet. Kiemelték, hogy bár számos bibliometriai munka foglalkozik az ezen a szinten végbemenő együttműködések elemzésével, a résztvevő entitások sokfélesége miatt (például az egyes országok eltérő kutatói létszáma, kutatásra fordított forrásai, kibocsátott publikációinak száma) nehéz érvényes méréseket végezni. Két ország (Svédország és Japán) együttműködéseinek keresztül mutatták be a mérési lehetőségeket. A tanulmányban a hálózatelemzés az ábrázolási lehetőségek bemutatásán keresztül jelent csak meg. A kutatás fókuszában az országok közötti kapcsolatok hasonlósági mutatókkal történő mérésének lehetőségei álltak. A szerzők elkülönítettek ún. kétoldalú (bilaterális) és többoldalú (mul-

tilaterális) hasonlósági mutatókat. A kétoldalú mutatókkal szemben, mint amilyen például a Salton- és a Jaccard-index, kritikát fogalmaztak meg, mivel azok mindig csak két ország egymáshoz való viszonyát írják le, valamint alábecsülik a kis országok, illetve egy kis és egy nagy ország közötti kapcsolat erősségét. Ezzel szemben a többoldalú mutatók a két ország viszonya mellett az egész vizsgált „világot” figyelembe veszik, összehasonlítva a várt és megfigyelt értékek arányát. A szerzők a Pearson-féle korreláció alkalmazhatóságát is vizsgálva megállapították, hogy mivel az két ország együttműködési mintázatának hasonlóságán alapszik, előfordul, hogy bár egyéb méréseknél gyenge kapcsolat áll fenn köztük, amennyiben mindkét ország ugyanazzal az ország-csoporttal működik együtt, akkor a köztük fennálló hasonlósági korreláció magas lesz.

Narin és munkatársai (1991) azt találták, hogy a nemzetközi társszerzőségek száma nem korrelál erősen egy adott ország tudományos tevékenységének nagyságával: adataik szerint a tudományos nagyhatalomnak számító Franciaországban vagy Kanadában kétszer-háromszor annyi nemzetközi társszerzős publikáció jön létre, mint a hasonló publikációs kibocsátású Japánban. Az adatok szórtságát Luukkonen és munkatársai (1992) is megállapították, ám épp ellenkező eredményre jutottak: értelmezésükben található negatív összefüggés az ország tudományos mérete és a nemzetközi társszerzőségek száma között. Ennek magyarázatát a tudomány specializálódásában keresték: az alacsonyabb tudományos tevékenységű országokban dolgozó kutatóknak nehezebb belföldön partnert találni a saját szakterületükön, valamint a költségek megosztására is nagyobb szükségük van.

Braun és munkatársai (1992) a nemzetközi társszerzős cikkeket vizsgálták a fizika aldiszciplínáin 1981 és 1985 között. Vizsgálatukban jelentős különbségeket mutattak ki a fizika altudományai között a nemzetközi együttműködések gyakorisága és erőssége szempontjából (pl. az akusztikában kevésbé, a részecskefizikában sokkal erősebbek a nemzetközi kapcsolatok). A fizika tíz alterületén mutatták be, hogy az egyes országok összes, ebbe a témába sorolt publikációja és ezen belül a nemzetközi társszerzős publikációk kapcsolata leírható egy hatványfüggvénnyel. Gyakorlatilag az együttműködések és publikációk száma közötti hatványfüggvényeloszlást írták le a skálafüggetlen hálózatok felfedezése előtt. A nemzetközi társszerzői hálózatot egy súlyozott irányítatlan hálózattal képezték le, melyben a kapcsolatok erősségét az országok közötti együttműködésekre számított Salton-indexszel határozták meg. Értekezésemben az egyes hálózatok meghatározásakor (részletesebben lásd a 6.1 és 6.2 fejezetben) én is a Salton-indexet használom a kapcsolatok erősségének számítására. Számomra ez azért is lényeges, mert így gyakorlatilag az egyes szereplők közötti kapcsolatok erősségét 0 és 1 közé normali-

zálom, és a társszerzőséget, valamint a kognitív távolságot leíró hálózatot ily módon tudom majd összehasonlítani.

Általánosságban Braun és munkatársai (1992) vizsgálati eredményei hasonlóak voltak a más tudományokhoz tartozó publikációkon mérteknél. Az Egyesült Államok és Nyugat-Európa központi pozíciója megkérdőjelezhetetlen volt: a maghoz tartozó országokat a skandináv országokkal Svédország, a keleti blokk államaival Lengyelország kötötte össze a vizsgált időszakban. A többi ország periférikus helyzetben állt a vizsgált tudományterületeken.

Wagner és Leydesdorff (2005) munkájukban szintén Salton-indexet használtak a nemzetközi társszerzői hálózatok kapcsolaterősségének leírására. Tanulmányukban ők nem egy tudományterületre szorítkoztak, sokkal inkább az 1990 és 2000 közötti időszak nemzetközi együttműködésekben bekövetkezett változásainak felderítését célozták meg, amikor is olyan események, folyamatok befolyásolták a nemzetközi kapcsolatok alakulását, mint a keleti blokk szétesése, az európai integráció vagy az internet elterjedése. A szerzők e változást háromféle módon elemezték: először feltérképezték az országok közötti társszerzős kapcsolatokat a globális régiók szintjén (EU-csatlakozó országok–Szovjetunió/FÁK; Amerika; szubszaharai Afrika; Ázsia–csendes-óceáni régió). Másodszor Braun és munkatársaihoz (1992) hasonlóan az országok kapcsolatának erősségét hasonlították össze a Salton-index segítségével. Harmadszor a társszerzői kapcsolatok szerkezetének, felépítésének felderítéséhez faktoranalízist végeztek, ahol az egyes országcsoportok alkották a faktorokat. A három módszer hasonló nemzetközi együttműködési mintázatokat adott eredményül: a globális kutatói hálózat növekedett, a kapcsolatok erősödtek, ugyanakkor új regionális központok is kialakultak (pl. a Dél-Afriai Köztársaság); a faktoranalízis arra világított rá, hogy a vezető tudományos nagyhatalmak (az Egyesült Államok, Nagy-Britannia és a Szovjetunió [1990] / Németország [2000]) „versengenek” egymással a kevésbé fejlett országok társszerzős kapcsolataiért, ugyanis ezek jellemzően csak egyiküket részesítik előnyben.

A (2) csoportba sorolt empirikus tanulmányok rávilágítottak arra, hogy az egyéb módszerek mellett hogyan lehet a hálózatelemzést is bevonni a társszerzőségek vizsgálatába. Az (1) csoportba sorolt tanulmányokban inkább a hálózatelemzési módszerek álltak a középpontban. Ezek a kutatások egyúttal arra is rámutattak, hogy a bibliometriai adatok milyen jó adatforrásai lehetnek a hálózatelemzésnek, és hogy az így létrejövő hálózatok milyen nagy (skálafüggetlen) hálózati modellbe illeszkednek. Ezzel szemben a (2) csoportba sorolt tanulmányokban a hálózatelemzés egy a többi vizsgálati módszer közül, konkrét kutatási kérdések megválaszolására, az eredmények alátámasztására szolgál.

4. Távolság értelmezése tudományos kapcsolathálózatokon

A tudományos együttműködések nem véletlenszerűen szerveződnek: létrejöttükben sok tényező játszik szerepet, ezek mentén mérhető a résztvevők közötti távolság. Tehát elmondható, hogy a távolság fogalmának segítségével mintegy számszerűsíthető két szereplő hasonlóságának mértéke. Egy bizonyos fokig minél jobban hasonlítanak egymásra az együttműködésben résztvevők, annál nagyobb eséllyel működnek együtt. Ugyanakkor a túl nagy hasonlóság, szélsőséges esetben az azonosság, ugyanúgy gátolhatja is az együttműködések létrejöttét, mint a túl nagy különbözőség: előbbi esetben az együttműködésben résztvevők ugyanazt tudják hozzáadni az adott projekthez, utóbbi esetben túl nagy energia-befektetést kíván a közös munka (Wuyts et al., 2005).

A távolság a regionális tudományi vizsgálatokban is lényeges térfogalom. A 4.1. fejezetben a távolságot és a hozzá köthető más térkategóriákat mutatom be. Ezt követően a bemutatott térkategóriák hálózatokon történő értelmezését vázolom. Ez a gondolati szint a területi kutatásokban egy újabb értelmezési réteget teremt, és a mögöttes ok-okozati tartalmak felkutatásában is segítséget jelenthet. Majd az együttműködések létrejöttének szempontjából lényeges tényezőket a szakirodalomban megismert távolságtípusokon keresztül mutatom be a 4.2. fejezetben. Végül a társadalmi és kognitív távolság viszonyát egy konceptuális modellben összegzem a 4.3. fejezetben.

4.1. Térkategóriák értelmezése hálózatokon

A regionális tudomány számára kiemelten fontos a térbeliség és az ahhoz köthető egyéb fogalmak, köztük a távolság vizsgálata, hiszen kutatásának gyújtópontjában a társadalmi és gazdasági folyamatok térrel összefüggésben történő vizsgálata áll (Nemes Nagy, 1998, 2003; Probáld, 2007; Rechnitzer, 2007).

A külső földrajzi térben megjelenő *távolság*, *szomszédság*, *helyzet*, *kitüntetett helyzetű fekvés*, *határmentiség* fogalmak a belső térben megvalósuló társadalmi és gazdasági folyamatokra is hatást gyakorolnak. Ezek számszerűsítésére és mérésére a regionális tudományban is újabban megjelenő regressziós módszerek szolgálnak, melyekben a térbeliség szerepe külön erős magyarázó jelleggel bír (Nemes Nagy, 2007). A vizsgált, távolsághoz köthető térkategóriákat és hálózatokon történő értelmezésüket egy korábbi tanulmányom alapján ismertetem (Vida, 2013).

Távolság

Távolság alatt leggyakrabban az euklideszi definíciót értjük, azaz két pont távolsága a köztük húzódó legrövidebb szakasz értékével határozható meg. Számszerűsítése a geometriában használt Pitagorasz-tétel segítségével történik (ld. pl. Nemes Nagy, 2009). Értelmezése és mérése sokféleképpen történhet. Alapvetően négy tényező mentén határozható meg, úgymint a vizsgálati egység, viszonyítási alap, távolságtípus és földrajzi lépték vagy területi keret (Nemes Nagy 2009; Fábián, 2012).

Két pont távolságának meghatározására az euklideszi távolságdefiníció mellett más módok is léteznek a szakirodalomban, mint például a Manhattan- vagy a Csebisev-távolság. Ezek más-más területen alkalmazhatók: míg az euklideszi távolságot például települések légvonalbeli távolságának meghatározásához használhatjuk, a Manhattan- vagy city-block távolság alkalmazása településen belül lehet hasznos. Ez a mód a manhattani utcahálózat négyzetrácsos voltára eredeztethető vissza, ahol a légvonalbeli és a valóságban megteendő út között nagy különbség adódott, így a befogók összege által meghatározott távolság jobban megfeleltethető volt a valóságnak. A Csebisev-távolság is a derékszögű háromszög befogóiból indul ki, de itt két pont távolsága a nagyobbik befogó hosszával egyenlő. Ezt jellemzően nem is a földrajzi térben használják távolság-meghatározásra, sokkal inkább belső térbeli jelenségek, fogalmak közötti különbség megállapítására szolgál, pl. vélemények, politikai pártok távolsága (Nemes Nagy, 2009).

A távolságot nem csak pontok, de különböző alakzatok, úgymint vonalak és poligonok között is meghatározhatjuk a kutatási kérdéstől függően, azonban a tényleges távolság értékének megadása ilyenkor oly módon történik, hogy a vonalat, illetve a poligont egy ponttal helyettesítjük valamilyen szempont alapján, és ezt követően már a két pont közötti távolság-meghatározások valamelyikét alkalmazhatjuk (Nemes Nagy, 2009).

A geometriai távolságon kívül más távolságértelmezések is használatosak: ilyen például az időtávolság, mely az egyik pontból a másikba történő eljutás idejét, vagy a költségtávolság, mely az utazásra fordított költséget veszi alapul a távolság számszerűsítésénél (Tóth és Kincses, 2007; Lőcsei és Szalkai, 2008; Dusek és Kotosz, 2016). Dusek és Kotosz (2016) *hálózati távolság* néven két pont között a közlekedési hálózatokon mért távolságot érti. Az elnevezés az értekezés szempontjából megtévesztőnek tűnhet, hiszen ezzel szemben én a hálózati távolságot általánosságban értelmezem a hálózatokon mint gráfokon.

A hálózaton belül a távolság tetszőleges két csúcs között az azokat összekötő kapcsolati lépések számával jellemezhető. Alapvetően két szinten vizsgálhatjuk a távolságot egy hálózaton belül: mikroszinten, két csúcs között, a leg-rövidebb úthosszal, makroszinten pedig, az egész hálózatra vonatkozóan, az átlagos úthosszal ragadható meg (ld. 2.3. fejezet).

A hálózatok egy része, az ún. területi hálózatok elemei rögzítettek a fizikai térben (pl. közlekedési vagy elektromos hálózatok), míg más hálózatok a fizikai térben nem rögzített elemek között írnak le kapcsolatot (pl. társadalmi, együttműködési hálózatok) (Dusek, Kotosz, 2016). Amennyiben egy hálózatot a fizikai térben is vizsgálunk, a hálózat felépítése, a csúcsok közötti kapcsolatok változatlanok maradnak, ugyanakkor a köztük lévő távolság immár kétféleképpen mérhető: hálózati értelmezésben a legrövidebb úthosszal, földrajzi értelmezésben jellemzően a fentebb ismertetett euklideszi távolsággal. Példaként a tömegközlekedési hálózatokat hozhatjuk. A budapesti 4-es metró vonalán Újbuda-központ egy-egy megállóra található a Bikás park és a Móricz Zsigmond körtér megállóhelyektől. Hálózati értelmezésben tehát a két távolság azonos, 1-1 lépés. A földrajzi térben azonban míg Újbuda-központ és a Móricz Zsigmond tér megállóhelyek között légvonalban (a Google Térkép adatai alapján) a távolság csak kb. 380 m, addig Újbuda-központ és a Bikás park megállóhelyek között kb. 1420 m. Láthatjuk tehát, hogy a fizikai és a hálózati topológiai távolság eltérő értékeket vehet fel.

Ugyanakkor az idő- és költségtávolság bizonyos elemei kifejezhetők a hálózati élek súlyozásával. Két szereplő (csúcs) között erősebb élsúlyt feltételezhetünk, ha például rövidebb elérési idő vagy kevesebb költség árán jön létre köztük kapcsolat. Ez természetesen nem független a fizikai térben kifejezett távolságtól.

Szomszédság

A szomszédság a távolsági helyzet egy speciális esete, leggyakrabban a poligonokkal (pl. régiók) kapcsolatban kerül előtérbe: amennyiben két poligon között a távolság 0, illetve van közös határuk, a két egység szomszédosnak tekinthető. Ilyenkor a köztük lévő kölcsönhatás erősebb, sőt a két egység tulajdonságaiban valószínűsíthetően hasonlít egymásra: ez egybevág Tobler „a földrajz első törvénye” néven ismert sejtésével (1970), mely szerint minden kölcsönhatásban van, de a közelebbi dolgok erősebben hatnak egymásra, mint a távolabbiak.

Különböző szomszédsági típusok (futár, vezér, bástya) különíthetők el aszerint, hogy hol húzódik a szomszédos alakzatok közös határa (Nemes Nagy,

1998, 2009; Jakobi és Jeney, 2008; Fábián, 2012). Alakzatok mellett pontok közötti szomszédság meghatározására is van mód: Jakobi és Jeney (2008) például nyolc lehetőséget mutatott be. Ezen lehetőségek közül hálózatokon a *minimális feszítőfa alapján meghatározott szomszédság* értelmezhető: eszerint „*i és j szomszédos, ha a minimális teljes élhosszú hurokmentes gráfban közvetlen éllel összekötött*” (Jakobi és Jeney, 2008, 297).

A szomszédság hálózatokon eszerint úgy értelmezhető, hogy köztük a lépések száma 1. Ezt elsőfokú szomszédságnak nevezzük. A hálózatok elemzése során fontos szerepet kap még a másodfokú (az elsőfokú szomszédok elsőfokú szomszédai) és a harmadfokú (a másodfokú szomszédok elsőfokú szomszédai) szomszédság is: ide köthető például Christakis és Fowler (2010) a 2.3. fejezetben említett három lépés hatótávolság-elmélete. A szomszédság a hálózatokon történő terjedés során kap kiemelt szerepet, hiszen csak az egymással szomszédos csúcsok között valósulhat meg az információ áramlása, de a járványok is hasonlóképpen terjednek (Newman, 2003; Barabási, 2016). A betegségek szomszédságon keresztül történő terjedését írja le például a SI, a SIR vagy a SIS modell (hálózati alkalmazásukhoz ld. pl. Newman, 2003).

Fekvés és helyzet

A fekvés a külső térben értelmezett helyzet, mely a térelemek egymáshoz való térbeli viszonyának feleltethető meg, a távolsággal és az iránnyal adható meg (Nemes Nagy, 2009). Tagai (2012) a fekvés tényezőit négy fő csoportra osztotta: földrajzi elhelyezkedés, centrumok elhelyezkedése, szomszédsági hatás és határmentiség. Ebből a centrumok elhelyezkedését és a határmentiség kérdéskörét a későbbiekben tárgyalom.

A hálózatok esetében a helyzet nem a külső térben értelmezett távolság és irány, hanem a csúcsok közötti kapcsolatok függvénye. Láttuk, hogy a távolság hálózatokon a lépések, azaz a két csúcs között húzódó köztes élek számával adható meg, az irány pedig csupán korlátozott formában jelenik meg: *A és B csúcs között a kapcsolat iránya mutathat A-ból B-be, B-ből A-ba, illetve oda-vissza, ez az ún. reciprok kapcsolat.* (Irányítatlan hálózatok esetén *A és B között egyenrangú a kapcsolat, irányról ilyenkor nem beszélhetünk.*) A hálózatban tehát az egyes csúcsok helyzetét a csúcs és a többi csúcs közötti kapcsolatok határozzák meg. Ez vizsgálható mikro- vagy lokális szinten, ilyenkor egy kiemelt csúcsot elemzünk annak kapcsolataival együtt (ez a csúcs ego-hálózata), illetve makro- vagy globális szinten, az egész hálózaton vizsgálva a csúcsok helyzetét.

Centrum és periféria

A centrum egyszerre foglalja magába a könnyű elérhetőséget a földrajzi térben, valamint a társadalmi-gazdasági térbeli kitüntetett helyzetet. Ebből az is következik, hogy egy térség centruma nem feltétlenül esik egybe annak középpontjával, mivel ez utóbbit tisztán geometriai tényezők határozzák meg (Lőcsei és Szalkai, 2008; Nemes Nagy, 2009; Tagai, 2012). A periféria felfogható a centrum ellentettjeként: csak egymáshoz viszonyítva léteznek. A külső térben ez a nehéz megközelíthetőséghez, a határmentiséghez kapcsolódik, míg a belső térben a centrumhoz képesti alárendeltségi viszonyhoz köthető (Nemes Nagy, 2009). A területi vizsgálatokban a centrum és a periféria ketőssége együttesen kerül feltárára, bemutatásra (Tóth, 2006; Tóth és Kincses, 2007; Lőcsei és Szalkai, 2008; Pénzes, 2013; Pénzes és Hegedűs, 2016).

A centrum-periféria duál hálózatokon is meghatározható. Ha egy hálózati megjelenítést látunk, az a benyomásunk támadhat, hogy a hálózat központi része az ábra közepén található, sok kapcsolattal rendelkező csúcsok együttese. Ennek oka, hogy a hálózatok vizualizációját meghatározó algoritmusok (pl. a Netdraw szoftver Springembedding vagy a Gephi szoftver Force Atlas és Yifan Yu algoritmusa) közös tulajdonsága, hogy jellemzően a több kapcsolattal rendelkező csúcsokat helyezik az ábra közepére, a periférikusabb helyzetű, kevesebb kapcsolattal rendelkező csúcsokat pedig a hálózat külső részén jelenítik meg. A megjelenítés alapján kialakuló intuáció a periféria esetében általában helyesnek bizonyul, ugyanakkor a centrum esetében a helyzet ennél árnyaltabb képet mutat.

A centrum és periféria lehatárolása a hálózatokban leggyakrabban a hálózat topológiája, struktúrája, a relacionális helyzeti viszony meghatározása alapján történik. Több hálózati szinten is értelmezhető (mikro-, mezo-, és makroszint). Mikroszinten a hálózat csúcsainak centralitásértékeivel ragadható meg (ezekhez lásd a 2.3. fejezetet), mezoszinten, irányított hálózatok esetén a diadikus kapcsolatok hierarchikus viszonyán keresztül vizsgálható (Wasserman és Faust, 1994), makroszinten pedig a hálózat egésze vagy egy összefüggő alhálózat a vizsgálat tárgya (Borgatti és Everett, 1999).

Makroszinten a centrum-periféria viszonyok Borgatti és Everett (1999) szerint három fő csoportba oszthatók:

- a hálózatban nem különíthető el egyértelműen centrum és periféria;
- a hálózatban egy sűrű központ és több, kisebb periférikus csoport alakul ki, melyek a központtal kapcsolatban állnak, egymással azonban nem. Ez a diszkrét modell, mely Freeman (1979) csillaggráfiának kiterjesztése;

- a hálózatban kialakul egy sűrű központ és egy azzal kapcsolatban lévő periféria, melyek között az átmenet a diszkrét modellel ellentétben nem éles, hanem folytonos.

A gyakorlatban a centrum-periféria duál meglétének igazolása a blokkmodellezés módszerével történhet (Batagelj, 1997; Doreian et al., 2004). A blokkmodellezés során a hálózati kapcsolatmátrix sorai és oszlopai – a hálózati relációk megváltoztatása nélkül – úgy kerülnek átrendezésre, hogy az egymással kapcsolatban álló csúcsok egymás mellé kerüljenek, blokkokat alkotva. Ezt követően az így kialakult kapcsolatmátrix a hálózattal megegyező méretű, a centrum-periféria relációt tisztán tartalmazó kapcsolatmátrixszal (8. ábra) kerül összehasonlításra, QAP korreláció segítségével.

	C	P
C	1	1
P	1	0

8. ábra: A centrum-periféria duál diszkrét modellje: *C* a centrum, *P* a periféria jele, a cellákban lévő 1 érték a kapcsolat létére, a 0 annak hiányára utal

Forrás: Borgatti és Everett, 1999 alapján saját szerkesztés

A folyamatos modellben a centrum-periféria átmenet jellemzéséhez minden csúcs egy „centrumszerűség” (*coreness*) értéket kap, a magas értékek a centrumra, az alacsonyak a perifériára jellemzők. Ezt követően ezen értékek csúcspáronként összehasonlításra kerülnek, így a centrum és a periféria mellett egy átmeneti, félperiférikus helyzet is létrejön.

A csúcsokhoz rendelt *coreness* mutatók a centralitásmutatókkal összehasonlíthatók. Borgatti és Everett (1999) felhívta a figyelmet arra, hogy míg a magas *coreness* értékkel rendelkező csúcsok jellemzően magas centralitáserteket vesznek fel, fordítva ez nem feltétlenül igaz. Jól példázza ezt, hogy a magas közteségcentralitással rendelkező csúcsok nem a hálózat magjához erősen kapcsolódó csúcsok, hanem azok, melyek a hálózat két vagy több, egymással kapcsolatban nem álló régióját kötik össze.

Határ

A határ fogalma alatt rendszerint olyan elemet értenek, mely két egységet választ el egymástól. Azonban szerepe valójában ennél összetettebb: Nemes Nagy (2009) négy jelentéstartalmat különített el: ezek az elválasztás, a szűrés, az ütközés és az összekapcsolás. A határfogalom hálózatokon is értelmezhető, mégpedig a hálózatok egyes moduljai között. A modul a hálózat csúcsainak egy olyan csoportja, melyben a csúcsok egymás között sok, viszont kifelé, a többi csúcs felé kevés kapcsolattal rendelkeznek (Csermely, 2005). (A hálózati modularitáshoz lásd részletesebben a 7.2.3. alfejezetet.) Ezen modulok különböző csoportkereső algoritmusok segítségével határozhatók meg. Az algoritmusok adta eredmények sokszor nem mutatnak teljes átfedést: az egyik módszer szerint az egyik, másik módszer szerint a másik modulhoz tartozó csúcsok tekinthetők egy ütköző „határzóna” részeinek is.

Azon szereplők, melyek több csoport felé is kevés kapcsolattal rendelkeznek, egyfelől megnehezítik (a kevés kapcsolat miatt) az információáramlást, másrészt lehetővé teszik azt, hiszen zárt modulokat kötnek össze egymással. Ezeket hídszereplőknek nevezzük (Csermely, 2005). Ez esetben a határ szűrő és összekötő szerepe is érvényesül (Nemes Nagy, 2009).

Földrajzi szélesség és hosszúság

A külső, földrajzi térben egy-egy pont egyértelműen megadható a földrajzi szélesség és hosszúság koordinátaival. Mivel így számértékeket rendelünk az egyes helyekhez, ezek térparaméterként vonhatók be a statisztikai vizsgálatokba (Nemes Nagy, 2009). Amennyiben a hálózat földrajzi tartalommal bír, lehetőség nyílik annak földrajzi térbe történő „kifeszítésére”: ekkor a hálózat csúcsai a földrajzi szélességi és hosszúsági koordinátákkal is megadhatók. Ezzel azonban a hálózatok alakja, megjelenítése jelentősen megváltozik, hiszen a csúcsok elhelyezkedését már nem az optimális kétdimenziós megjelenítés, hanem a koordinátákkal kifejezett földrajzi helyzetük határozza meg. A hálózat jellemzésekor a csúcsokhoz rendelt koordináták a csúcsok tulajdonságaiként foghatók fel. Így vizsgálhatók például a hálózatok sűrűségének területi egyenlőtlenségei.

Az 1. táblázat a főbb regionális tudományi térkategóriák hálózati értelmezését foglalja össze. Láthatjuk, hogy a térkapcsolatok leképezhetők a hálózatok szintjén is, bár eltérő, néhol szűkebb értelmezésben: ennek oka, hogy a hálózatok a valóságot a kapcsolatokon keresztül képezik le, az egyéb jelentéstartalmak ennek alárendelve jelennek meg.

Térkategória	Értelmezés
távolság	két csúcs között megtett út alapján az érintett élek száma, azaz a lépések száma
irány	irányított hálózatok esetében háromféle lehetőség létezhet két csúcs között. A csúcsból B-be, B-ből A-ba és a két csúcs között oda-vissza mutató kapcsolat.
szomszédság	egy lépés távolságra levő, közvetlen élen elérhető csúcsok, minimális feszítőfa alapján meghatározott szomszéd
centrum	mikroszinten magas centralitásértékekkel rendelkező csúcsok, makroszinten magas coreness értékkel rendelkező csúcsok; magasabb sűrűséggel rendelkező csoportok a hálózaton belül
periféria	mikroszinten alacsony centralitásértékekkel rendelkező csúcsok, makroszinten alacsony coreness értékkel rendelkező csúcsok; alacsonyabb sűrűséggel rendelkező csoportok a hálózaton belül; <i>potenciálisan a rendszerhez tartozó, de bekapcsolatlan csúcsok</i>
határ	csoportokat összekötő hídszereplők; <i>a hálózat megszakíthatósága, részekre szakadása</i>

1. táblázat: Térkategóriák értelmezése hálózatokon

Forrás: Vida, 2013, dőlttel Nemes Nagy, 2017 kiegészítései

4.2. Távolságtípusok a tudományos együttműködésekben

Az eddigiekben áttekintettem a távolság és az ahhoz köthető regionális tudományi térkategóriák hálózatokon történő értelmezését. A kutatói hálózatok felfoghatók a belső tér elemei között fennálló viszonyrendszerként is, melyekben a résztvevők közötti távolság a kapcsolatok létrejöttében, megvalósulásában fontos szerepet játszik.

Doktori kutatásomban a szerzők közötti hasonlóságot vagy különbözőséget a távolságon keresztül azonosítom, hiszen minél jobban hasonlít egymásra két szereplő, annál kisebb közöttük a távolság. A hasonlóság az előző fejezetben ismertetett térkategóriák közül kifejezhető a szomszédsággal, illetve a közelséggel. A szereplők közötti távolságot egy-egy, az együttműködések létrejöttét meghatározó tényező, illetve távolságtípus mentén vizsgálom, és így tulajdonképpen egy-egy távolságtípus egy-egy dimenzióként fogható fel. A szakirodalomban számos ilyen tényezőt különböztetnek meg: a francia *Dynamique de proximité* (Távolságdinamika) kutatócsoport tett először különbséget a gazdaságtudományban a fizikai térben meglévő földrajzi távolság, valamint az együttműködésekben résztvevők kapcsolati, tudásbeli terének átfedéseit mutató szervezeti távolság között (lásd pl. Torre és Gilly, 1999 összefoglalását). A szakirodalomban sokan hivatkoznak Boschma (2005) ötös felosztására:

- kognitív távolság: a tudásbázis hasonlósága, illetve különbözősége
- szervezeti távolság: az együttműködők hierarchiában betöltött szerepének, autonómiájának, illetve kötöttségének vizsgálata
- társadalmi távolság: a résztvevők közötti személyes kapcsolatok, bizalom
- intézményi távolság: a résztvevő intézmények (a tudományos együttműködések esetében általában az egyetemek, piaci szereplők és kormány) mechanizmusai, a tudásmegosztás szabályai és szokásai közötti különbségek
- földrajzi távolság: a szereplők közötti fizikai térbeli távolság

Saját kutatásom szempontjából célszerűbbnek tartottam egy hármas tipológiát, melyben Boschma (2005) szervezeti, társadalmi és intézményi távolságát egyetlen társadalmi távolságtípusba soroltam: dolgozatomban tehát földrajzi, kognitív, illetve társadalmi távolságtípus-csoportok között teszek különbséget. Ennek oka abban áll, hogy a társadalmi távolság – a többi távolságtípushoz hasonlóan – sokféleképpen mérhető; én ebből egyet, a társ-szerzői kapcsolatok vizsgálatát választottam ki, így kutatásom nem terjed ki az együttműködésekben résztvevők egy bizonyos szervezetben egymáshoz képest betöltött helyére, sem pedig arra, vajon az együttműködők ugyanaból a szektorból érkeztek-e.

A következőkben áttekintem a dolgozatban elkülönített három távolságtípust, a témában végzett korábbi kutatásokkal együtt. Bemutatom, hogy a szakirodalomban milyen módszerekkel mérték a kérdéses távolságokat, illetve főbb eredményeiket is ismertetni fogom.

4.2.1. Földrajzi távolság

Az első távolságtípus, a földrajzi távolság az, melyet a valós, földi térben értelmezhetünk (Nemes Nagy, 2009): ez a szereplők földrajzi térben meghatározott helyei között mért távolság (Zitt et al., 2000; Boschma, 2005; Frenken et al., 2009a; Hoekman et al., 2010; Acosta et al., 2011). A tudománymetria földrajzi távolsággal foglalkozó ágát Frenken és munkatársai (2009a) „térbeli tudománymetria” (*spatial scientometrics*) néven nevezték, és vizsgálatának tárgyát a szakirodalom alapján három pontban határozták meg: a tudomány térbeli eloszlása, térbeli irányultságok (például belföldi együttműködések előnyben részesítése a nemzetközivel szemben), idézettség. Ebből főleg a másodikkal fogunk foglalkozni, de az idézettség térbeli alakulására is láthatunk példát.

A távolság mérhető kilométerben, légvonalban (pl. Katz, 1994), de elérési időben is: Ponds és munkatársai (2007) a holland kutatóintézeteket vizsgáló tanulmányukban például az autós és a tömegközlekedési idők súlyozott átlagával dolgoztak. A szakirodalomban több mérési szintet különítenek el: a távolság mérhető városok (pl. Katz, 1994 az egy városon belüli kutatóintézetek között önkényes távolság-meghatározással), régiók vagy megyék (pl. Acosta et al., 2011 az EU NUTS3 szintjén), de országok között is (pl. Nomaler et al., 2013). E két utóbbi esetben fokozottan merül fel a kérdés, hogy a régiók, illetve országok mely pontjai között mérendő a távolság: Acosta és munkatársai (2011) a régiók fővárosait választották vonatkoztatási pontnak, míg Nomaler és munkatársai (2013) a két ország legnagyobb városai közötti bilaterális távolságok súlyozott átlagát használták, ahol a súlyozás aszerint történt, hogy az adott város az ország népességének hány százalékát adja. Láthatjuk, hogy ez utóbbi esetben a földrajzi távolság egy társadalmi dimenzióval, a népességgel is bővül.

Katz (1994) az elsők között vizsgálta a földrajzi távolság szerepét a kutatói együttműködések kialakulásában. Az Egyesült Államok, Nagy-Britannia és Ausztrália belföldi egyetemközi kapcsolatait vizsgálva arra a következtetésre jutott, hogy a távolság növekedésével az együttműködések száma exponenciális csökkenést mutat. Katz ezt azzal magyarázta, hogy a kutatói együttműködések feltétele az informális, személyes kapcsolat (Boschma, 2005 tipológiájában az alacsony társadalmi távolság), melynek kialakítása, fenntartása nagy földrajzi távolságban nehéz. Hasonló eredményre jutottak például Frenken és munkatársai (2009a), valamint Hoekman és munkatársai (2009) is európai uniós régiók közötti tudományos együttműködések vizsgálatakor. Frenken és munkatársai (2009b) megállapítása szerint ez azért különösen érdekes, mert napjaink tudományos életében a fapados repülőjáratok térhódításával, az internet elterjedésével, valamint az angol közvetítőnyelvvé válásával egyenesen „a távolság haláláról” szokás beszélni (pl. O’Brien, 1992; Cairncross, 1995, 2001). Mások (pl. de Blij, 2012) amellet foglaltak állást, hogy a földrajzi távolság szerepe egyáltalán nem elhanyagolható napjainkban sem.

A földrajzi távolság és a kutatói együttműködések fent vázolt, egyszerűnek tűnő kapcsolatához Ponds és munkatársai (2007) szolgáltak érdekes adalékokkal, igaz, az eddig tárgyalt cikkeknel földrajzilag jóval kisebb alapterületről. Holland egyetemközi és egyetemi-ipari együttműködések vizsgálata során arra a következtetésre jutottak, hogy a nagyobb földrajzi távolság csak a szektorok közötti együttműködést befolyásolja negatívan, az egyetemközit nem. Ezt azzal magyarázták, hogy az egyetemi-ipari együttműködések során a Boschma (2005) által definiált intézményi távolság eleve nagyobb, így az alacsony földrajzi távolság ezt „kompenzálja” az együttműködés kialakulása-
kor.

A földrajzi távolság idézettségére gyakorolt hatását országok szintjén először Narin és munkatársai (1991) mutatták be: eredményeik szerint a nemzetközi társszerzős cikkek idézettsége mindegy kétszerese a belföldi társszerzőségben elkészült publikációkéénál. A távolság nagyságának szerepét Nomalerésmunkatársai(2009) egyelméleti modellfelállításával bizonyították. Tanulmányukban EU-n belüli együttműködések vizsgálata során arra az eredményre jutottak, hogy minél nagyobb a távolság két együttműködő ország között, annál magasabb lesz a közös publikáció idézettsége. A földrajzi távolság vizsgálata az értekezésben a társadalmi és kognitív kapcsolatok térképre vetítésén keresztül kerül elemzésre, így vonva be a külső földrajzi teret a vizsgálatba.

4.2.2. Társadalmi távolság

A földrajzi távolsággal ellentétben ez egy olyan távolságtípus, mely a belső térben értelmezendő, azaz a földrajzi tértől elválasztva rendelkezik saját viszonyrendszerrel, elrendezéssel (Nemes Nagy, 2009). A társadalmi távolsággal foglalkozó kutatások a jelenséget rendszerint a szereplők közötti társadalmi-gazdasági tényezők hasonlóságán (pl. Narin et al. 1991; Zitt et al. 2000; Frenken et al., 2009a; Acosta et al. 2011), illetve a társszerzőségek, együttműködések kialakulásának szociális dimenzióján keresztül (pl. White et al., 2004) vizsgálták. A társadalmi dimenzióra hatással van az egyén környezete: az együttműködések kialakulásánál döntő tényező, hogy kik a kollégái, kiket ismer, kikkel találkozott egy-egy konferencián, milyen intézményi struktúrában dolgozik, milyen nyelven publikál stb. A társadalmi távolság vizsgálatához a bibliometriai eszközökön túl (társszerzőségek vizsgálata) gyakran használnak más módszereket is (pl. interjú, kérdőív) (pl. White et al., 2004; Autant-Bernard et al., 2007). Doktori kutatásomban én a társadalmi távolságot a társszerzőségen keresztül vizsgálom.

A társadalmi távolságot először a közgazdaságtanban vizsgálták: Granovetter (1985) a gazdasági tevékenység társadalmi beágyazottságáról beszélt, melyben a társadalmi kapcsolatok ugyanakkor meg is határozzák a gazdasági tevékenység eredményét. Boschma (2005) értelmezésében ugyanakkor a társadalmi távolságban egyfajta „arany középút” a célravezető: az innovációt, az új tudás létrehozását mind a túl magas, mind a túl alacsony társadalmi távolság negatívan befolyásolhatja. A következőkben a társadalmi távolság főbb szegmenseit ismertetem, a vonatkozó szakirodalom csoportosításával, főbb eredményeinek bemutatásával.

Az együttműködésekben részt vevő kutatók társadalmi távolságánál első helyen említhetjük a *belföldi / nemzetközi társszerzőség* kérdését. Sokat kutatott tény, hogy a kutatói együttműködésekben a belföldi társszerzőségek száma

messze felülmúlja a nemzetköziet (ld. pl. Narin et al., 1991; Luukoonen et al., 1993; Hoekman et al., 2010): Nomaler és munkatársai (2003) ezt például európai uniós régiók szintjén is kimutatták. A jelenség magyarázatában a korábban tárgyalt földrajzi közelség, illetve a meglévő személyes, informális kapcsolatok mellett a közös háttértudás (közös nyelv, kulturális tényezők, a tudomány működésének helyi mechanizmusai) is nagy szerepet játszik (Frenken et al., 2009a, Hoekman et al., 2010). Ugyanakkor Hoekman és munkatársai (2010) azt találták 33 európai ország 2000–2007 közötti tudományos együttműködéseinek vizsgálatánál, hogy a nemzeti, illetve regionális határok elválasztó szerepe csökkenő tendenciát mutat, helyét egyre inkább a földrajzi távolság veszi át.

Összekapcsolódik a belföldi partnerek előnyben részesítésével a *nyelvi-kulturális közelség* kérdése is. Annak ellenére, hogy az angol egyre inkább a tudomány „közös nyelveként” funkcionál (ld. pl. Siguan, 2001), a nemzetközi együttműködéseknel nagyobb a kutatók hajlandósága arra, hogy azonos, de legalábbis hasonló anyanyelvű partnereket válasszanak: Liang és munkatársai (2006) negatív exponenciális kapcsolatot mutattak ki az EU-15 tagállamok körében az országpárok közötti Salton-index és a (nyelvcsaládok ágai szintjén értelmezett) nyelvi távolság között. Hasonló jelenségre hozott példát Csomós (2017, 2018) is: európai egyetemek legfontosabb külföldi partnerországait (az Egyesült Államokat nem számítva) vizsgálva azt találta, hogy sok esetben a választás nyelvileg-kulturálisan meghatározott. Svájcban például a német, francia, illetve olasz nyelvű területen fekvő egyetemek kutatói rendre Németországgal, Franciaországgal és Olaszországgal működnek együtt leginkább. Hasonló a helyzet a kétnyelvű (francia–flamand) Belgiumban is. A kelet-közép-európai államok legfontosabb partnere a történelmi-kulturális tényezők miatt Németország, kivéve Szlovákiát és Romániát: előbbi a közös történelem miatt Csehországot, utóbbi pedig nyelvi-kulturális hatások miatt Franciaországot részesíti előnyben.

A *gazdasági közelség* kérdését járták körül például Acosta és munkatársai (2011) az EU-15 régiói szintjén. Kutatásukban arra az eredményre jutottak, hogy két régió között minél nagyobb az egy főre jutó kutatás-fejlesztésre jutó kiadások különbsége, annál kisebb a tudományos együttműködés valószínűsége. Ez magyarázható azzal, hogy a jobb helyzetben lévő régióknak nem érdekük az alacsonyabb K+F ráfordítású régiókkal az együttműködés. Amint arra a szerzők rámutattak, az eredmény ellentmond a centrum-periféria elméletnek (a tudományos együttműködések szintjén ld. pl. Schott, 1998), mely szerint az alacsonyabb költségvetésű régiók kutatói hajlamosabbak a „gazdagabb” régiókból együttműködő partnert találni. A jelenség magyarázata az lehet (Acosta et al., 2011), hogy az EU-15-ös régiók között nem elég nagy a gazdasági távolság ahhoz, hogy a centrum-periféria modell érvényes

legyen. A centrum-periféria modell érvényességét Wagner és Leydesdorff (2005) is megkérdőjelezték, ugyanakkor Acosta és munkatársai (2011) magyarázatával ellentétben egy olyan modellt állítottak fel, melyben több nagy gazdasági-tudományos teljesítményű ország „verseng” a kisebb potenciálú országokkal való együttműködésért.

A szereplők közötti társadalmi és gazdasági tényezők hasonlóságát az *intézményi távolságon* keresztül is lehet vizsgálni, mely fogalom alatt Boschma (2005) definícióját követve a különböző típusú intézmények (egyetemek, cégek, kormányzati szervek) tudományos együttműködése értendő, a felek folyamatos párbeszédével, kölcsönhatásával – ezt a jelenséget írja le Leydesdorff és Etzkowitz (1998) „hármasspirál” modellje is. Ponds (2009) nyolc technológiai jellegű tudományágban vizsgálta a különböző típusú együttműködések arányát 1988 és 2004 között. Nem meglepő módon, az akadémiai szervezetek fedték le a közös publikációban megvalósuló együttműködések kb. 50%-át: ezt az akadémia-kormány, majd az akadémia-gazdaság kapcsolatok követték. A különböző szektorok között nemzetközi együttműködés ritkábban valósult meg, összhangban Ponds és munkatársai (2007) azon megállapításával, hogy az alacsonyabb mértékű földrajzi távolság „kompenzálja” a nagyobb szektoriális különbséget. Blumenthal és munkatársai (1996) a cégek oldaláról vizsgálták az akadémiai szervezetekkel való együttműködést, kitérve annak problémáira is, mint például a szellemi tulajdon felhasználásának kérdése: az egyetemeknek az új tudás terjesztése is érdeke, szemben a cégek azon törekvésével, hogy a szellemi tulajdon védve legyen, más ne férhessen hozzá (ld. pl. Hall, 2003).

Két kutató *szervezeti távolságát* az határozza meg, hogy egy adott intézmény hierarchiáján belül milyen kapcsolatban állnak egymással (Boschma, 2005). Egy korai, a kérdést egy egyetemi kampuszon vizsgáló tanulmányukban Kraut és munkatársai (1988) a szervezeti (kutatócsoportok, tanszékek, laboratóriumok) mellett a fizikai távolságot is figyelembe vették (épületszárny, emelet, épület). Egyik, némileg triviális eredményük, hogy azok a kutatók, akiknek irodái egymáshoz közel helyezkedtek el, többször vettek részt együttműködésben, és az irodák közötti távolság növekedésével az együttműködések aránya csökkent. Ez annak fényében, hogy az irodák elrendezése rendszerint az intézmény szervezeti struktúráját követi, egyértelműnek tűnik.

A szervezeti távolság kapcsán Kretschmer (1994) szolgáltatott érdekes adalékkal: intézményen belüli együttműködések vizsgálva arra az eredményre jutott, hogy azok a kutatók működnek együtt gyakrabban, akiknek a publikációs aktivitása között nagyobb a különbség. Ebből az következik, hogy egy kevés publikációval rendelkező, kezdő kutató nagyobb valószínűséggel választja a sokat publikáló, tapasztaltabb, idősebb kutatót, mint a magához hasonló

pályakezdőt: ez egyfajta mester-tanítvány viszonyt feltételez. A „nagy tudomány” (de Solla Price, 1979) tipikus kutatócsoportjainak felépítése is ezt követi: a csoport egy vezető kutató felügyeletével működik, tagjai közé frissen doktorált kutatók, doktoranduszok, de egyetemi hallgatók is tartoznak.

Egy, szintén a kérdést célzó kutatásban Liang és munkatársai (2000) különböző életkorú, számítógép-tudománnyal foglalkozó kínai kutatók együttműködését vizsgálták. Eredményeik azt mutatták, hogy kiemelkedő a fiatalabb és idősebb kutatói korosztály együttműködése, akár két fiatal – egy idős felállásban is. (Nem mellékes körülmény, hogy a vizsgált időszakban az 1960-as, 70-es évek „kulturális forradalma” miatt a fiatal középkorú kutatók száma rendkívül alacsony volt.) Az, hogy ezekben az esetekben rendszerint az idősebb professzor a második (harmadik) szerző, a kínai kultúrában gyökerező segítő mester képével magyarázható.

A társadalmi távolság mérésében a *személyes kapcsolatok* is nagy szerepet kapnak. Több kutatás (pl. Crane, 1972; Maglaughlin és Sonnenwald, 2005) is kiemelte a kapcsolati háló fontosságát a jövőbeli kutatói együttműködések kialakításakor. A kollegiális kapcsolatok arra is használhatók, hogy általuk új, lehetséges partnerekkel lépjen egy kutató kapcsolatba: Newman (2001a) hálózattudományi vizsgálatának egyik eredménye volt, hogy két szerző között megnő az együttműködés valószínűsége, ha van egy közös társszerzőjük. A társadalmi hálózatok vizsgálatában ezt „a barátom barátja a barátom” jelenséget nevezik tranzitivitásnak (Wasserman és Faust, 1994).

Más mutatókkal együtt a tranzitivitást is vizsgálták Autant-Bernard és munkatársai (2007) európai uniós kutatás-fejlesztési együttműködések hálózatain: kutatásukban olyan mikro- és nanotechnológiával foglalkozó cégeket vizsgáltak, akik közös projektben vettek részt az Európai Unió FP6-os keretprogramjában. Az együttműködések kialakulásának valószínűségét háromféle szempontból vizsgálták: az egyes vállalkozások jellemzői (kutatói háttér, büdzsé), hálózati hatások (két cég közötti távolság a kapcsolati hálózaton, az egyes cégek közvetlen és két lépés távolságra lévő partnereinek száma) és földrajzi hatások (a két cég székhelye közötti távolság kilométerben kifejezve, belföldi/nemzetközi együttműködés, az adott országok centrum-periféria helyzete). Regressziós modelljük eredményei szerint nemzetközi szinten a (hálózattudományi módszerekkel értelmezett) társadalmi távolság jobban előre jelezte az együttműködések kialakulását, mint a földrajzi távolság: összhangban az eddigiekkel, ha i és j , valamint j és k cég között volt már kialakult kapcsolat korábbi együttműködésekben, az megnövelte annak esélyét, hogy i és k cég közös projektet adjon be.

A személyes kapcsolati háló kapcsán ki kell térni az így formálódó „láthatatlan kollégiumok” jelenségére. De Solla Price (1979) a „nagy tudomány” egyik sajátosságaként beszélt olyan kutatói közösségekről, melyek tagjai ismerik és felhasználják egymás munkásságát, illetve nemzetközi pályázatok segítségével gyakori látogatásokat tesznek egymás kutatóintézetében. Crane (1972) ezen csoportosulások jellemzőit egy kiterjedt, matematikus és szociológus kutatók bevonásával végzett vizsgálatában, bibliometriai eszközökkel és interjúkkal tárta fel. Definíciója szerint a láthatatlan kollégium egy kutatási területet vizsgáló kutatók alcsoportjának kommunikációs hálózata (Crane, 1972). Fontos jellegzetessége, hogy nem intézményhez kötődik, alapja a kutatók közötti társas kapcsolat. A kommunikáción keresztül a kutatók célja egy közös probléma megoldása, ezt segítik az informális beszélgetések, egymás munkásságának nyomon követése, az így kialakuló együttműködések. A tudományos probléma megoldásakor a láthatatlan kollégium „feloszlik”.

Ugyanakkor a láthatatlan kollégium modellje többek szerint nem tartható: Laki és Palló (2001) összefoglalójukban az internet adta megnövekedett kommunikációs lehetőségek, ez által pedig a kutatói hálózat méretének nagy arányú növekedése miatt tartották idejétmúltnak a láthatatlan kollégium fogalmát, Lievrouw és munkatársai (1987) pedig a kutatói együttműködések társadalmi és kognitív vonatkozásainak különbségére hívták fel a figyelmet.

4.2.3. Kognitív távolság

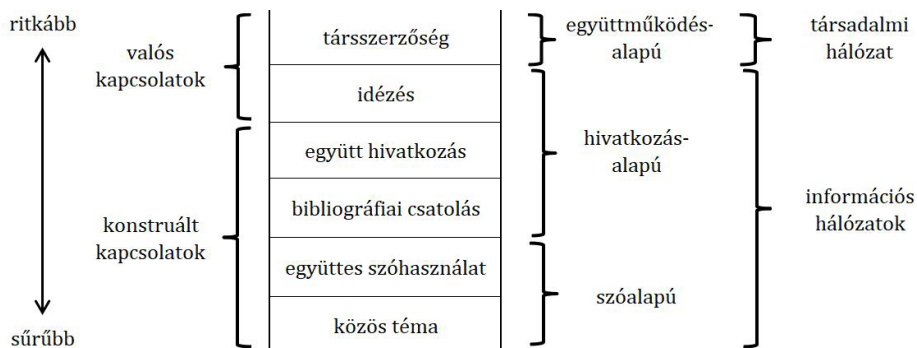
A kognitív távolság alatt a kutatók tudásbázisa közötti különbséget értjük (Kessler, 1963; Small, 1973; Jarneving, 2007; Frenken et al., 2009a; Boyack és Klavans, 2010; Yan és Ding, 2012). A társadalmi távolsághoz hasonlóan, ez a távolságtípus is a belső térben értelmezhető (Nemes Nagy, 2009). Az eddigiekben láttuk, hogy olyan tényezők, mint a fizikai közelség, illetve a társadalmi-személyes kapcsolatok mennyiben befolyásolják az együttműködések létrejöttét. Ettől nem teljesen független, ám mégis más alapokon áll ez a mérési mód: amennyiben két fél tudása túlságosan különböző alapokon nyugszik, túl sok erőfeszítésbe kerül számukra az új tudás kommunikálása, megértése, valamint feldolgozása (Boschma, 2005). Más szavakkal, mivel a (gazdasági, tudományos) szereplők mindegyike saját kutatási profillal, érdeklődési körrel rendelkezik, a tudásbázis átfedésének mértéke határozza meg a kognitív távolságot.

Bármilyen előnyösnek tűnhet az alacsony kognitív távolság, nem minden esetben hasznos: a vonatkozó szakirodalom alapján Boschma (2005) három érvet fejtett ki egy nagyobb távolság mellett. Először, az új tudás létrehozását általában segíti, ha a két fél meglévő tudása kiegészíti egymást, mivel a különböző nézőpontok segítenek az új, kreatív ötletek megtalálásában. Másodszer,

a túl kicsi távolság egyfajta kognitív „bezárkózást” eredményezhet, ahol a feladatok rutinszerű megoldása elfedi az innováció lehetőségét. Végül, a közelség megnöveli az akaratlan tudásmegosztás kockázatát, hiszen a tudás nem sajátítható teljesen ki: így emiatt az egymáshoz hasonló területen dolgozó szereplők általában vonakodnak tudásukat megosztani egymással.

Az optimális kognitív távolság „fordított U” modelljét (ld. pl. Nooteboom et al., 2007) Wuyts és munkatársai (2005) piaci szereplőkön, két kísérlettel igazolták. Elméletük szerint az optimális távolság feltétele, hogy megmaradjon mind a kölcsönös érthetőség, mind az újdonságélmény előnye. Az első vizsgálatban gyógyszerészeti vállalatok és kisebb biotechnológiai cégek közötti kutatás-fejlesztési együttműködések szerepeltek. A kognitív távolságot indirekt módon, az együttműködések számával mérték: kiinduló hipotézisük szerint a sorozatos együttműködés két szereplő között lecsökkenti a tudásbázis különbségeit. Az együttműködések sikerességét azok a szabadalmak jelezték, melyeket a vizsgált időintervallumban az Amerikai Gyógyszerügynökség (FDA) új hatóanyagként jegyzett be. A kutatás eredményei szerint azok a gyógyszercégek voltak a legsikeresebbek (azok nyújtották be a legtöbb innovatív szabadalmat), melyek sem túl kevés (túl alacsony kognitív távolság), sem túl sok (a tudás „szétszabdálása”) partnercéggel működtek együtt. A második vizsgálat mikrotechnológiai cégek adatain alapult: itt a kognitív távolság egy alkotórészeként a cégek szervezeti felépítésének hasonlósága jelent meg. A szervezeti felépítések optimális különbözősége erősebb faktornak bizonyult az együttműködések sikerességének meghatározásánál, mint a technológiai, tudásalapú hasonlóság mértéke. Ebből a példából is láthatjuk a kognitív távolság és a társadalmi távolság összefonódását: az együttműködések száma vagy a szervezeti felépítés hatással volt a tudásmegosztásra.

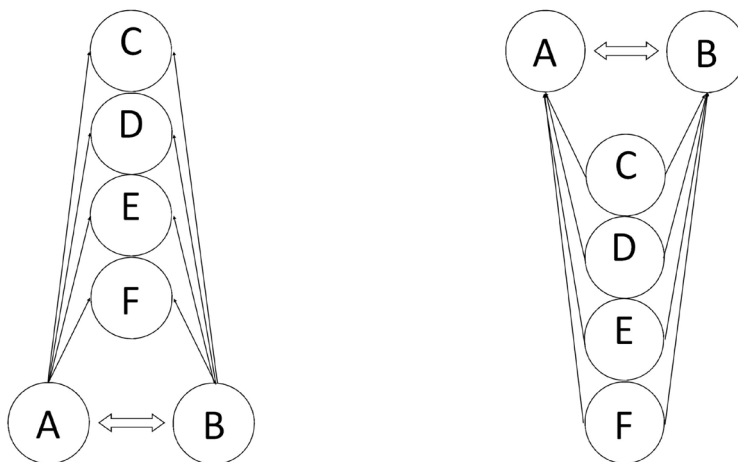
Az eddigiekből is kiderült, hogy a kognitív távolság mérése az előző két távolságtípusénál nehezebb: itt nem áll rendelkezésünkre a fizikai tér vagy a társszerzőségekkel megfogható kapcsolatok rendszere, mint a földrajzi és a társadalmi távolságnál. Azt is mondhatjuk, hogy a kutatói együttműködési hálózatok egy részében valójában mesterségesen létrehozott kapcsolatokon keresztül mérjük a távolságot. Yan és Ding (2012) a 9. ábrán szemléletesen foglalta össze a kutatói együttműködési hálózatok vizsgálatához leggyakrabban alkalmazott módszereket, azok jellegzetességeivel együtt. Fontos kiemelni, hogy munkájukban a valós földrajzi térbeli kapcsolatokat nem vették figyelembe, így a kutatói együttműködések a társadalmi és kognitív (információs) kapcsolatok mentén vizsgálták, melyek közül a társszerzőséget a társadalmi távolság terén már említettem.



9. ábra: Kutatói együttműködési hálózatok vizsgálatának módszerei

Forrás: Yan és Ding, 2012 alapján saját szerkesztés

A különböző mérési módszereket részletesebben a fejezet további részében tárgyalom, egyelőre csupán röviden definiálom őket. Az *idézés* során úgy valósul meg kapcsolat két szereplő között, hogy az egyik a másik munkájára hivatkozik. Az *együtt hivatkozás* (Small, 1973) és a *bibliográfiai csatolás* (Kessler, 1963) is hivatkozásból indul ki: az előbbinél két szereplő között akkor van kapcsolat, ha egy harmadik szereplő publikációjában mindkettőjüket idézi, míg az utóbbi esetében akkor, ha két szereplő közösen hivatkozik egy harmadik szereplő munkájára. A két módszer egymáshoz való viszonyát ábrázolja a 10. ábra.



Bibliográfiai csatolás

Együtt hivatkozás

10. ábra: Bibliográfiai csatolás és együtt hivatkozás: a körök publikációkat jelölnek, a fekete nyilak a hivatkozás irányát. Az A és B között fennálló kapcsolatot jelöli az átlát-szó kétirányú nyíl

Forrás: Garfield, 1988 alapján saját szerkesztés

Amint azt a 10. ábrán láthatjuk, a két módszer nagyon hasonlít egymásra: az ábra bal oldalán A és B publikáció bibliográfiai csatolás eredményeként kapcsolatban áll egymással, mivel mindkettő a C–F betűkkel jelölt dokumentumokra hivatkozik. Az együtt hivatkozás esetén a folyamat fordított: A és B között azért jön létre kapcsolat, mivel C–F publikációk mindegyike hivatkozik rájuk. A bemutatott módszerek közül ez utóbbi az egyetlen, mely nem az adott egyén tevékenysége, hanem az őrá hivatkozók alapján szerveződik. A módszertani eltérések következtében a bibliográfiai csatolás és az együtt hivatkozás segítségével különböző információk azonosíthatók: míg az előbbi egy diszciplína kurrens kutatási irányának felmérésére alkalmas, az utóbbi arra mutat rá, mi az a „közös alap”, melyből egy tudományág művelői kiindulnak (Boyack és Klavans, 2010).

Az *együttes szóhasználat* (Callon et al., 1983) akkor köt össze két szereplőt, ha az általuk írt publikációkban (címben, absztraktban stb.) ugyanazok a kifejezések jelennek meg. Ennek általánosítása a *közös téma* (Tang et al., 2008): a közösen előforduló szavak gyakoriságából következtetni lehet arra, vajon két szereplő ugyanabban a kutatási témában működik-e.

Yan és Ding (2012) felosztásukban először egy skálán helyezték el a vizsgálati módszereket a belőlük kialakuló hálózatok sűrűsége szerint (9. ábra). Vizsgálatukból az az általános tendencia rajzolódik ki, hogy a társszerzőségi hálózatok sűrűsége alacsonyabb, mint a más, hasonlóságon alapuló mérési módszerekből meghatározott hálózatoké. Ezt azzal magyarázták, hogy a társszerzőség olyan valós társas kapcsolat, mely nehezebben alakul ki, mint a hasonlóságon alapuló megközelítések, vagy akár a szintén valós kapcsolatként definiált közvetlen idézés. Ezzel szemben az együtt hivatkozást, a bibliográfiai csatolást, az együttes szóhasználatot, illetve a közös témákat felderítő eljárások mesterséges, konstruált kapcsolatot eredményeznek. Ezeknél a hálózatoknál a kapcsolat a hasonlóságon keresztül keletkezik két szerző között, a két kutatónak nem kell feltétlenül ismernie egymást ahhoz, hogy hasonló témákkal foglalkozzanak, tehát él legyen köztük a hálózatban. (Gondoljunk csak a múltban kis időbeli eltéréssel feltalált találmányokra.)

A 9. ábra jobb oldala bemutatja, hogy az egyes módszereknek milyen típusú adatok képezik az alapját. Az idézés, együtt hivatkozás és a bibliográfiai csatolás kiinduló hipotézise az, hogy a publikációk háttérében lévő tudásbázis megismerhető az adott publikáció hivatkozásainak (vagy a publikációra történő hivatkozások) elemzésén keresztül. Az együttes szóhasználat és a közös téma a szövegekben előforduló azonos kifejezések együttállása alapján állapítja meg, vajon két dokumentum mögött hasonló tudás áll-e. A társszerzőség mint módszer az együttműködéseket vizsgálja: ebből a nézőpontból két kutató közötti együttműködés arra utal, hogy azonos témát kutatnak.

A társadalmi és információs hálózati felosztás mögött az áll, hogy a társszerzőség megvalósulásához a személyek valamilyen fokú közvetlen kapcsolata szükséges, míg a többi módszer által a szerzők között teremtett kapcsolathoz nincs szükség közvetlen kapcsolatra, elégséges a szükséges információkat tartalmazó publikáció elérése.

A különböző vizsgálati módszerek céljaikban is eltérnek egymástól. A közvetlen idézési kapcsolatokból kialakuló hálózat jól mutatja a tudás áramlását, az új felfedezések elterjedését. A különböző konstruált kapcsolatok a különböző tudományágak, tudományos trendek azonosítására, magának a tudománynak a feltérképezésére alkalmasak (Yan és Ding, 2012).

Bibliográfiai csatolás

A bibliográfiai csatolás módszerének megalkotása Kessler (1963) nevéhez fűződik, célja az volt, hogy a közösen hivatkozott publikációk rendszerén keresztül olyan koherens, egy-egy tématerületnek megfeleltethető csoportokat körvonalazzon, melyek aztán az információgyűjtésben nyújthatnak segítséget (Jarneving, 2007). A módszer leggyakrabban témadetekcióra használatos, ekkor lépései a következők (Boyack és Klavans, 2010 szerint):

- a közösen hivatkozott publikációk számának felhasználásával a hasonlóság kiszámítása a mintában lévő valamennyi publikáció között páronként;
- a mintában szereplő publikációk klaszterekbe rendezése a hasonlósági értékek alapján (erre a lépésre csak akkor van szükség, ha tématerületet szeretnénk azonosítani a módszer segítségével).

A tématerületek feltérképezésén kívül a szakirodalomban más példákat is találunk a bibliográfiai csatolás alkalmazására. Glänzel és Czerwon (1996) például a módszer segítségével azonosította az egyes tudományterületek „törzsdokumentumait” (*core documents*), melyek segítségével a szerzők szerint az egyes tudományágakban aktuális, friss témák deríthetők fel. A törzsdokumentumok a bibliográfiai csatolással létrehozott hálózatokban központi helyre kerülnek, sok és viszonylag erős kapcsolatuk révén. Bár az együtt hivatkozást gyakrabban alkalmazzák a szakirodalomban kognitív hálózatok létrehozásához, Glänzel és Czerwon (1996) megállapítása szerint az újonnan felmerülő, kialakuló kutatási témák felmérésére a bibliográfiai csatolás módszere alkalmasabb, mivel időbe kerül, hogy egy új témában annyi publikáció szülessen, melyen érvényes együtt hivatkozási vizsgálatot lehet végezni. Az egyes országok szintjén vizsgálva a törzsdokumentumokat, azt találták a szerzők, hogy a nagyobb tudományos teljesítménnyel rendelkező országok (USA, Japán, nyugat-európai országok) által kibocsátott törzsdokumentumok száma és aránya is magasabb az átlagnál. A legtöbb törzsdokumentum az

élettudományok (különböző orvostudományi témák) területén született, bár az is igaz, hogy az ilyen témájú publikációk eleve felülreprezentáltak voltak a mintában, mivel alapja a SCI volt.

A bibliográfiai csatolás egy másik alkalmazását találhatjuk Boyack (2008) tanulmányában. Vizsgálatában először a folyóiratok, majd a szerzők szintjén hozott létre bibliográfiai csatoláson alapuló klasztereket, ezután ezeket elhelyezte egy tudománytérképen. A klaszterekhez affiliációkat rendelt, majd ezt követően azt mérte fel, hogy két választott intézmény (egy amerikai tudományegyetem és egy kutatólaboratórium, szintén az Egyesült Államokból) mely tématerületeken publikál, illetve mekkora a vitalitásuk ezeken a területeken. A vitalitást az 5 évnél nem régebbi publikációkra történő hivatkozások arányával mérte. Végül olyan potenciális együttműködési területeket vázolt fel, melyeken mindkét intézmény részéről születtek publikációk, és mindkét intézmény esetében magas vitalitásértéket jelzett a mutató.

Ebből a példából is láthatjuk, hogy a bibliográfiai csatolás nem csak a tudományometriai elemzések alapegysége, a publikációk, hanem más szintek, azaz szerzők, intézmények, folyóiratok stb. között is alkalmazható. Szerzői szintre történő kiterjesztése különösen fontos lesz a következő fejezetben, melyben az ez irányú korábbi kutatások (Zhao és Strotmann, 2008a, 2008b) ismertetése után saját konceptuális modellemet vázolom fel a szerzői bibliográfiai csatolás kapcsán.

Együtt hivatkozás

Small (1973) nevéhez köthető az a másik, hivatkozásokon alapuló eljárás, mely két dokumentum kapcsolatát a mindkettőre egyszerre hivatkozó publikációk száma alapján határozza meg. A módszer témadetekcióra történő alkalmazásakor a következő (Boyack és Klavans, 2010):

- a közös hivatkozások száma alapján a hasonlóság kiszámítása a mintába tartozó publikációk között páronként;
- a publikációk klaszterekbe sorolása a hasonlósági értékek alapján;
- a hivatkozó publikációk hozzárendelése a hivatkozott dokumentumok klasztereihez.

Boyack és Klavans (2010) itt különbséget tesz együtt hivatkozáson alapuló klaszterezés (*co-citation clustering*) és együtt hivatkozások elemzése (*co-citation analysis*) között: míg az előbbinél csak a hivatkozott dokumentumok kerülnek a klaszterekbe, az utóbbi a hasonlósági értékek alapján a hivatkozó publikációkat is (egyben vagy parciálisan) egy klaszterhez rendeli. Ahogy

a módszer leírásából is látszik, a szerzők a második megoldás mellett döntöttek vizsgálatukban. Természetesen, ahogy a bibliográfiai csatolásnál is, a klaszterekbe sorolás itt sem képezi a módszer magját: a leírás azt a gyakorlatot tükrözi, hogy az együtt hivatkozás leggyakoribb alkalmazása homogén tematikus csoportok, klaszterek létrehozása.

Az együtt hivatkozás előnye Small (1973) érvelése szerint, hogy időben változó mintázatok vizsgálatára is alkalmas lehet, mivel az előzőekben tárgyalt bibliográfiai csatolással ellentétben nem egy lezárt, hanem egy folyamatosan alakuló folyamat vizsgálatára alkalmas. Míg a bibliográfiai csatolás egy fix, állandó kapcsolat a csatolt dokumentumok hivatkozásain keresztül, az együtt hivatkozást befolyásoló hivatkozó publikációk száma egyre növekszik, így az idézési mintázatok is változhatnak.

A módszer témadetekciós és dinamikus lehetőségeinek alkalmazására jó példa Braam és munkatársai (1991a, 1991b) két tanulmánya. Az első (1991a) egy olyan módszert körvonalazott, melyben az együtt hivatkozás kulcsszóelemzéssel párosult, bár ez utóbbi valójában eszköz volt az együtt hivatkozás eredményeinek validálására. Mezőgazdasági kémiai tárgyú publikációk elemzése során a szerzők kimutatták, hogy az együtt hivatkozással létrejött klaszterek koherens, egymástól különböző tématerületeket fednek le, bár igaz, hogy egy-egy tématerület „szétszóródhat” több klaszterre. Ugyanakkor azt is megjegyezték, hogy a választott tudományág kurrens munkái csupán részben azonosíthatók a módszer segítségével, a szerzők eltérő hivatkozási gyakorlatai miatt.

Braam és munkatársai az előző folytatásának is tekinthető tanulmányukban (1991b) atom- és molekuláris fizikai témájú publikációkon elemezték az együtt hivatkozásokból, valamint a kulcsszavakból kialakuló klaszterek időbeli változását egy tíz éves időtartamon keresztül. Elemzésükben arra az eredményre jutottak, hogy a kulcsszavakból kirajzolódó időszerű kutatási témák kevésbé változtak a vizsgált időintervallumban, mint a kiinduló alapnak számító szakirodalom. Ez utóbbi annak is betudható a szerzők szerint, hogy az ugyanazt a témát kutató szerzők több „iskolára” oszlottak, külön intellektuális alapnak számító publikációkkal.

Az együtt hivatkozás társadalmi vetületére hívták fel a figyelmet White és munkatársai (2004). Egy 16 főből álló interdiszciplináris kutatócsoport vizsgálatában arra keresték a választ, vajon a hivatkozási mintázatok tükrözik-e a csoport társas szerkezetét. Az egymásra hivatkozások száma nőtt a kutatócsoport működése folyamán, a baráti és kollegiális kapcsolatok megnövezték az egymásra hivatkozás esélyét: ez a hivatkozás társadalmi dimenzióját támasztja alá. Ugyanakkor a szerzők a kutatás legfontosabb eredményeként azt

emelték ki, hogy az eddig felsorolt faktoroknál az együtt hivatkozás jobban előrejelzi az egymásra hivatkozást: a mintában részt vevő kutatók hajlamosabbak voltak azt idézni, akivel közösen idézve voltak egy másik publikációban. Ez az eset tehát – összhangban a láthatatlan kollégiumok kritikájával – a kognitív távolság erősebb voltát támasztja alá a társadalmi távolsággal szemben. Igaz ugyanakkor, hogy a kis létszámú minta nem tesz lehetővé túlzott általánosítást.

Idézés

Az idézés (az angol nyelvű szakirodalom *direct citation*, *inter-citation*, illetve *cross-citation* néven ismeri) az eddig részletesebben tárgyalt módszerekkel ellentétben valós kapcsolatot: két szereplő között akkor áll fenn, ha egyikük idézi a másikat. Amint láthatjuk, ilyenkor a két szereplő, az idéző és az idézett között irányított kapcsolat áll fenn, így a létrejövő mátrix aszimmetrikus lesz. Hogy itt is lehessen koszinusz hasonlóságot mérni, előbb szimmetrikussá kell tenni a mátrixot az adatok tükrözésével (Boyack és Klavans, 2010). Fontos még kiemelni, hogy az idézést jellemzően folyóiratok vagy szerzői csoportok szintjére aggregált dokumentumokon vizsgálják (Janssens et al., 2009).

Az idézés jellegzetes alkalmazása a szakirodalomban a tudománytérképek létrehozása (pl. Boyack et al., 2005; Leydesdorff és Rafols, 2009; Carley et al., 2017). A tudomány egészének feltérképezése a kutatás-fejlesztés szempontjából kiemelten fontos: a résztvevőknek ismerniük kell saját tudományáguk közvetlen környezetét, ezenkívül az egyes tudományágak publikációiban mért aktivitása, illetve a „térképen” elfoglalt helyzete az anyagi támogatásokat érintő döntések meghozatalára is hatással lehet (Boyack et al., 2005). Tudománytérképeken az egyes intézmények (egyetemek, kutatóintézetek, ipari szereplők) tevékenységi profilja is ábrázolható (Carley et al., 2017), amint azt az előbbi, bibliográfiai csatolást tárgyaló részben is láthattuk (Boyack, 2008). Boyack és munkatársai (2005) különböző, idézésből és együtt hivatkozásból kiinduló hasonlósági mutatók segítségével próbálták „a tudomány gerincét” feltérképezni. Tanulmányukban nyolc mutató összehasonlítása során arra az eredményre jutottak, hogy az idézésen alapuló hasonlósági értékekből létrejövő klaszterek megbízhatóbban fogják össze az azonos témakörbe tartozó folyóiratokat, mint az együtt hivatkozásokon alapuló. A klaszterek hálózati ábrázolásával a szerzők a központi és periférikus tudományágakat mutatták be: minél magasabb egy csúcs fokszáma, annál több más területtel van kapcsolata. Ebből a szempontból a vizsgált mintán a biokémia bizonyult a „leginterdiszciplinárisabb” tudományágnak.

Együttes szóhasználat, közös téma

Míg az eddig tárgyalt három módszer a publikációk hivatkozásából kiindulva igyekezett az egységek (publikációk, szerzők, folyóiratok, intézmények stb.) közötti kognitív távolságot meghatározni, az együttes szóhasználaton és közös témán alapuló vizsgálatok magában a cikkek szövegében (leggyakrabban a címben vagy az absztraktban) előforduló kifejezések alapján feltételeznek kapcsolatot (Callon et al., 1983). A kifejezések lehetnek még szakértők által megadott kulcsszavak is (Ding et al., 2001). A módszer Ding és munkatársai (2001) alapján a következőképpen írható le:

- a megfelelő szövegrészletben (cím, absztrakt stb.) kulcsszavak gépi vagy manuális kiemelése;
- a kifejezések normalizálása (ragozott alakok, szinonimák kiszűrése);
- mátrix létrehozása, melyben a sorokban és oszlopokban az azonosított kifejezések szerepelnek, a cellaértékben pedig az, hogy a két kifejezés hány alkalommal fordult elő azonos szövegben;
- az együttes előfordulási mátrix hasonlósági mátrixszá alakítása hasonlósági mutatók segítségével;
- a kapott értékek segítségével a kifejezések klaszterekbe sorolása.

A módszer alkalmazására az eddigiekben is láthattunk már példát (pl. Bram et al., 1991a, 1991b), ám itt az együttes szóhasználat vizsgálata csak az alkalmazott hivatkozásalapú módszer validálására szolgált, annak ellenőrzésére, hogy tematikusan valóban koherens csoportok jöttek-e létre a klaszteranalízis után. A módszer önálló alkalmazására jó példa Ding és munkatársai (2001) vizsgálata: együttes szóhasználatot alkalmazó elemzésükben az információtudomány gyakori témáinak változását követték nyomon egy tíz éves periódus alatt. A talált kulcsszavakat öt klaszterre bontották, ám vizsgálatukból kiderül, hogy ezek a klaszterek a vártnál kevésbé koherensek, illetve nem is határolódnak el élesen egymástól. A tíz éves időtartam két öt éves szakaszra bontásával a szerzők az információtudomány témáinak változását kívánták illusztrálni: az eredmények azt mutatják, hogy a tudományágnak vannak bizonyos állandó témái, ugyanakkor az új témák is gyorsan betagozódnak a rendszerbe.

Módszerek összehasonlítása, hibrid módszerek alkalmazása

A szakirodalomban több tanulmány (pl. Boyack és Klavans, 2010; Yan és Ding, 2012) arra irányult, hogy az eddig tárgyalt módszerekből kapott eredmények összehasonlításával, illetve ezek alapján új, hibrid módszerek kialakításával minél pontosabb képet kaphassanak a tudomány különböző szintjeiről. Az eddigiekben láthattuk, hogy az egyes módszerek kiindulópontja különbözik egymástól, ennek megfelelően bizonyos jelenségek mérésére egyik módszer alkalmasabbnak bizonyul a másikinál.

Yan és Ding (2012) hat módszert hasonlított össze kutatói hálózatok létrehozásában: társszerzőség, idézés, bibliográfiai csatolás, együtt hivatkozás, együttes szóhasználat, közös téma. A kapott hálózatokat Salton-féle koszinusz hasonlósággal egybevetve arra a következtetésre jutottak, hogy a legnagyobb hasonlóság a bibliográfiai csatolással és közvetlen idézéssel létrejött hálózatok között mutatkozott, míg a legerősebb különbözőséget a társszerzői és közös témán alapuló hálózatok mutatták.

A hat módszer hasonlóságait a szerzők többdimenziós skálázás segítségével két dimenzióban ábrázolták: az első dimenzió felfogható a „nem hivatkozás-alapú – hivatkozás-alapú” dichotómia ábrázolásaként, a második pedig a „társadalmi – kognitív” vetületként. Ezt követően, a kutatói együttműködések minél teljesebb vizsgálatához különböző hibrid eljárásokat javasoltak: az első kettőben az első, illetve a második dimenzió tengelyén egymástól legtávolabb eső hálózatok (1. dimenzió: közös téma – együtt hivatkozás; 2. dimenzió: társszerzőség – együttes szóhasználat) kombinálását mutatták be, a harmadik módszerben pedig valamennyi hálózat egyetlen hasonlósági mátrixban történő egyesítését.

Yan és Ding (2012) célja tehát nem minőségi sorrend felállítása volt a kognitív távolságot mérő módszerek között, hanem az általuk kapott eredmények kombinálásával új, pontosabb módszerek felállítása. Más tanulmányokban arra találunk példát, hogy egy létrehozott hibrid módszert hasonlítanak össze más módszerekkel. Janssens és munkatársai (2006, 2009) az idézés és az együttes szóhasználat alkalmazásával kialakult klasztereket hasonlították össze egy, a két eljárás összevonásával létrejött csoportosítással. Eredményeik szerint mind egy tudományágat, az információtudományt (Janssens et al., 2006), mind pedig Web of Science-adatok alapján a tudomány egészét (Janssens et al., 2009) tekintve, a hibrid módszer koherensebb, egymástól jobban elkülönülő klasztereket hozott létre, mint a két módszer külön-külön alkalmazása. Janssens és munkatársai (2009) a kapott klasztereket egy létező sémával, a 22 tudományterületet azonosító ESI (*Essential Science Indicators*) tipológiával is összehasonlították: eredményeik szerint az általuk javasolt

hibrid eljárással koherensebb csoportok jöttek létre, és bár egyes klaszterek csaknem teljes egészében megfeleltek egy-egy ESI kategóriának (pl. a kémia esetében), az átfedés nem volt teljes.

Boyack és Klavans (2010) különböző gyakran használt elemzési módszereket (idézés, bibliográfiai csatolás, együtt idézés) hasonlított össze, valamint negyedik módszerként egy bibliográfiai csatoláson és együttes szóhasználaton alapuló hibrid eljárást is vizsgáltak. Biokémiai tárgyú publikációkon végzett kutatásukban kiemelték, hogy a témadetektálásra, a koherens klaszterek létrehozására a „hagyományos” módszerek közül a bibliográfiai csatolás a legalkalmasabb, ám ennél is pontosabb képet ad a hibrid eljárás. Kutatásuk érdekessége, hogy a klaszterek koherenciájának vizsgálatába bevonták a publikációk azon adatait is, hogy milyen támogatott projektből valósultak meg: hipotézisük szerint az egy projektből megvalósuló publikációk szoros tematikus kapcsolatot mutatnak, tehát azonos klaszterbe kell kerülniük. Az általuk megalkotott hibrid megoldás itt is valamivel jobb eredményt adott, mint a többi módszer.

4.3. Társadalmi és kognitív távolság viszonya a szerzői bibliográfiai csatoláson keresztül – egy konceptuális modell

Az előző fejezetben bemutatam az együttműködések létrejöttét befolyásoló, az azokban résztvevő szereplők közötti, a disszertációban elkülönített három távolságtípust. Ezek közül a belső térben jelenlevő társadalmi és kognitív távolság nem függetlenek egymástól. A többek között gazdasági, kulturális, ismeretségi és nyelvi tényezők elősegítik az együttműködések létrejöttét és gyakran kapcsolatban állnak a kognitív tényezőkkel is. A társadalmi távolság mentén egymáshoz közel álló szereplők kognitív távolsága is gyakran kicsi, hiszen a hasonló gazdasági és kulturális háttér sokszor hasonló tudásbázis is von maga után. Ugyanakkor a kognitív térben egymáshoz közel álló szerzők a társadalmi dimenzió mentén is közel lehetnek egymáshoz. Gondoljunk csak például arra, hogy a hasonló tudással és érdeklődési körrel rendelkező kutatók nagyobb eséllyel ismerik egymást, mivel munkájuk során több lehetőségük van a találkozásra (pl. kollégák, közös konferencián vesznek részt stb.).

Doktori kutatásomban a társadalmi távolságot az előző fejezetben felvázolt lehetőségek közül a társszerzőségek révén azonosítom, míg a kognitív távolság mérését az elérhető módszerek közül a bibliográfiai csatolással végzem.

A tudományometriai elemzések alapegységei a publikációk. A létrehozott tudás leginkább (de nem kizárólag) publikációkon keresztül kerül a tudományos közösség számára hozzáférhetővé. Ez az a szint, melyre vonatkozóan közvetlenül érhetők el az adatok a bibliometriai és citációs adatbázisokból.

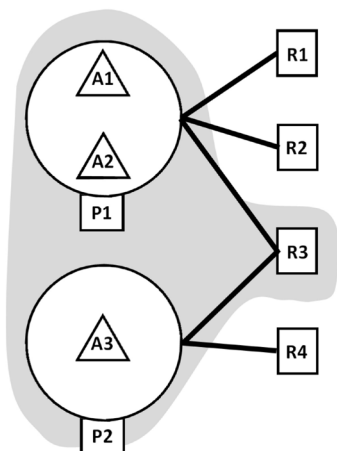
A publikációs rekordok természetesen számos olyan információt tartalmaznak, melyek lehetővé teszik a publikációk szintjénél magasabb és alacsonyabb aggregáltsági szinten is a vizsgálatokat. Ilyen magasabb aggregáltsági szint pl. a folyóiratok szintje, míg alacsonyabb szintnek a szerzők szintje tekinthető. A szerzők affiliációs adatainak segítségével is mód van magasabb aggregáltsági szintű elemzéseket végezni, úgymint pl. az egyes intézmények, települések, régiók, országok publikációs kibocsátása. A magasabb szinteken mért adatokra vonatkozó elemzések megbízhatóbbak (Bornmann et al., 2008; Bornmann és Marx, 2013), egyrészt a számosság, másrészt az adatbázisokban szereplő adatok megbízhatósága miatt, illetve a magasabb szintre vonatkozó elnevezések pontosabbak is. Például egy ország nevének az elírása ritka, és könnyebben kiszűrhető az elemzés során, mint amikor ugyanez egy szerzővel történik.

A tudományos együttműködések vizsgálatának alapszintje a szerzői szint. Doktori kutatásomban a vizsgálathoz én is ezt a szintet választottam. A szerzők affiliációi révén ebben az esetben is van lehetőség magasabb szinteken vizsgálni az együttműködések, pl. intézmények, települések, megyék, régiók, országok között (Yan és Ding, 2012; Vida 2016).

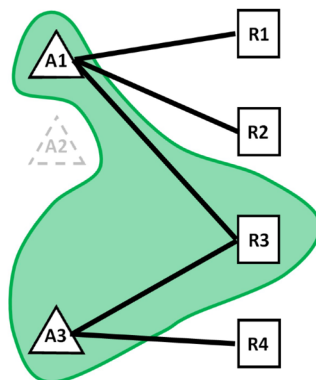
A 4.2. fejezetben bemutattam a bibliográfiai csatolás módszerét, mint a kognitív távolság mérésének egyik lehetőségét. A módszert leggyakrabban téma-területek meghatározására használják a publikációk szintjén, mérve az egyes publikációpárok közötti hasonlóságot a közös hivatkozások alapján. Minél magasabb a bibliográfiai csatolás értéke két szerző között, annál nagyobb átfedést mutatnak az általuk felhasznált szakirodalmi tételek, azaz kognitív távolságuk annál kisebb.

A 11. ábra a bibliográfiai csatolás publikációk szintjéről a szerzői szintre történő vetítésének módját mutatja, egyrészt Zhao és Strotmann (2008a, 2008b) munkája, illetve doktori kutatásomban történő alkalmazása alapján. Az ábra a társadalmi és a kognitív távolság kapcsolatát is szemlélteti a szerzői bibliográfiai csatolás esetén. P a publikációkat, A a szerzőket és R a hivatkozásokat jelöli.

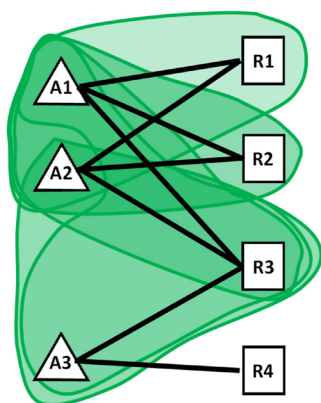
A 11. ábra bal felső része a hagyományos bibliográfiai csatolás módszerét mutatja a publikációk szintjén. $P1$ publikációnak két szerzője és három hivatkozása, míg $P2$ publikációnak egy szerzője és egy hivatkozása van. A $P1$ és $P2$ cikk egymással bibliográfiai csatolást mutat $R3$ közös hivatkozáson keresztül.



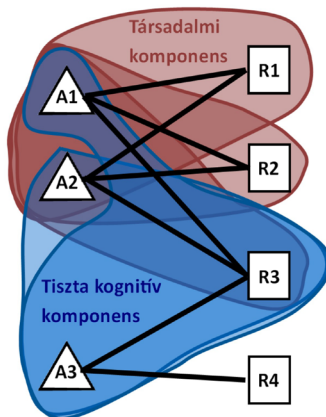
Bibliográfiai csatolás



Szerzői bibliográfiai csatolás
Zhao és Strotmann (2008a, 2008b) alapján



Módosított szerzői bibliográfiai csatolás



**Módosított szerzői bibliográfiai csatolás
komponensekre bontása**

11. ábra: Bibliográfiai csatolás szerzői szintre vetítése

P: publikáció, A: szerző, R: hivatkozás

Amennyiben nem a publikációk, hanem a szerzők között vizsgáljuk ugyanezt, a bibliográfiai csatolás szerzői szintre történő kiterjesztését kell alkalmazni, ahogy ezt például Zhao és Strotmann (2008a, 2008b) tette. A szerzői szintre vetítéskor ők azonban csak az első szerzőket vették figyelembe, azaz a publikációk hivatkozásait csupán az első szerzőkhöz rendelték hozzá (11. ábra jobb felső része). Az ábrán látható, hogy a *P1* publikáció *A2* szerzője a szerzői bibliográfiai csatoláskor nem jelenik meg a mintában, tehát a kognitív távolság ezzel a módszerrel csak első szerzők között mérhető.

Doktori kutatásomban a szerzői szintre vetítéskor én nem alkalmaztam ilyen kitételezt, hanem Radicchi és munkatársai (2009) szerzői együtt hivatkozás vizsgálatakor alkalmazott, minden szerzőre kiterjedő vetítését alkalmaztam. Az adott publikáció összes hivatkozását hozzárendeltem a cikk minden egyes szerzőjéhez, így az összes szerzőt be tudtam vonni a vizsgálatba (11. ábra bal alsó része). *P1* publikáció mindkét szerzőjéhez hozzárendelésre került a *P1* publikáció mindhárom hivatkozása. *P2* publikáció egyszerűs volt, így ott az *A3* szerzőhöz kerültek hozzárendelésre a *P2* cikk hivatkozásai.

Az összes szerző figyelembevételéért is lényeges, mert egy szerző a publikációinak csak egy részében első szerző, így a Zhao és Strotmann (2008a, 2008b) által alkalmazott módszertan szerint munkásságának egy része rejtve marad. A szerzői szintű bibliográfiai csatolásnál a szerzők között nézzük a közös hivatkozásokat. Zhao és Strotmann (2008a, 2008b) vetítési módjakor szerzői bibliográfiai csatolás áll fenn *A1* és *A3* szerzők között az *R3* hivatkozáson keresztül (11. ábra jobb felső része). Az általam alkalmazott vetítés során *A1* és *A2* között valósul meg szerzői bibliográfiai csatolás *R1*, *R2*, *R3* hivatkozáson keresztül, továbbá *A1* és *A3*, valamint *A2* és *A3* is szerzői bibliográfiai csatolásban van egymással *R3* hivatkozáson keresztül (11. ábra bal alsó része). Vegyük észre, hogy *A1* és *A2* szerzői bibliográfiai csatolása a társszerzőség miatt alakult ki. Ezt a komponenst *társadalmi komponensnek* neveztem el, mivel ez a társszerzőségből következett. Ezzel szemben *A3* az *A1* és *A2* szerzőkkel külön-külön is szerzői bibliográfiai csatolást mutat *R3* hivatkozáson keresztül: ez a szerzők kognitív kapcsolata miatt van, így ezt *tiszta kognitív komponensnek* neveztem el (11. ábra jobb alsó része).

Az általam módosított szerzői szintű bibliográfiai csatolás azt eredményezte, hogy a kapott kognitív távolság – ezt a doktori értekezésben *teljes kognitív távolságnak* neveztem – még további két távolságtípust is tartalmazott, nevezetesen a *társadalmi távolságot* és a *tiszta kognitív távolságot*. A komponensek segítségével jól látható a kognitív és a társadalmi távolság kapcsolata. Az összes szerzőre történő vetítés egyúttal magával hozta azt, hogy a társszerzőségek léte is befolyással bírt a szerzők közötti kognitív távolság mértékére. Tulajdonképpen a társszerzőség, mellyel a társadalmi távolságot azonosítottam,

hatással van a teljes kognitív távolságra. Az értekezésben elkülönített tiszta kognitív távolság, bár mentes a társszerzőségekből adódó hasonlóságtól, hatást gyakorolhat a társadalmi távolságra a jövőbeli együttműködések kijelölése révén. A tiszta kognitív távolság segítségével azonosítani lehet a jövőbeli potenciális együttműködők körét, hiszen ők azok, akik ugyanazzal a tématerülettel foglalkoznak, de még nem társszerzők. Természetesen az értekezés empirikus részében, mivel a vizsgálat csak öt évre terjedt ki, előfordulhat, hogy a tiszta kognitív kategóriában azonosított szerzők a vizsgálati időszakot megelőzően már publikáltak közösen.

5. Bibliometriai adatok elérése, kezelése – a vizsgálat adatbázisa

Az értekezés további részében azonosítom és összehasonlítom a konceptuális modellben bemutatott három távolságtípust: a társadalmi, a teljes és a tiszta kognitív távolságot. Bebizonyítom, hogy indokolt a teljes kognitív távolság két komponensre bontása, mivel a tiszta kognitív és a társadalmi távolság empirikus adatok alapján is elkülöníthető egymástól. Ebben a fejezetben ismertetem a kutatás adatbázisát, illetve hogy a tudományometriában milyen adatbázisok érhetők el, ezek használata milyen előnnyel és hátránnyal jár. A kutatáshoz használt WoS adatbázist részletesebben is bemutatom. Ezt követően a 6. fejezet tárgya az, hogy a meghatározott távolságtípusok milyen módon számszerűsíthetők, az így kapott értékek többféle módszerrel történő összehasonlítását pedig a 7. fejezet tartalmazza.

5.1. Az „ideális” adatbázis

Ideális adatbázis nem létezik, hiszen ha létezne, mindenki csak azt használná. A helyette létező és elérhető adatbázis adatait így kénytelen a kutató olyan formába hozni, hogy felhasználhassa a vizsgálatában: például meg kell tisztítani az abban található neveket, hogy egy bizonyos dolognak az elnevezése mindig azonos formában forduljon elő, és így már egyértelműen azonosítható legyen stb.

Az adatbázisok egy-egy dologra vonatkozóan sok adatot tárolnak rendezett formában (www.kfki.hu/~cheminfo/hun/eloado/adatb/adatb2.html). Ez lehetővé teszi az adatok vizsgálatát és így módon az új információhoz való hozzájutást.

Milyen lenne mégis az ideális adatbázis, ha létezne? A vizsgálati kérdés megválaszolásához szükséges adatok – a választott módszerhez illeszkedően – egy helyen, tisztított és sztenderdizált formában, azonnal valamilyen adatelemző vagy statisztikai szoftverbe betölthető módon lennének elérhetők. Ez a leírás két fő részre osztható: az egyik a vizsgálati kérdéshez szükséges adatokat, azaz az alapkorpust érinti, míg a másik az adatok minőségét (tisztított, széleskörű, jól kereshető, megfelelő további munkához alkalmas formátumban elérhető stb.).

Egy ideális adatbázisban az alapkörpusz mindig teljes körű lenne, és lefedné teljesen a vizsgálni kívánt populációt. Ehhez fontos a vizsgálati kérdés megfelelő konceptualizálása és az ehhez történő alapkörpusz meghatározása. Egy ideális adatbázistól azt várhatnánk, hogy az elméleti alapkörpusz meghatározásával megegyező méretű és minőségű adatok állnak rendelkezésre benne.

Ez a valóságban nincs így. A kutatók az adataikat elsődleges vagy másodlagos forrásokból érik el. Elsődleges vagy primer adatok esetén a kutató maga veszi fel az adatokat, ehhez számos mintavételezési eljárás és elv ismert (pl. Babbie, 2004; Vicsek, 2006). Másodlagos vagy szekunder adatok esetén a kutató már valamilyen mások által gyűjtött és létrehozott adatbázis adatait használja fel. Mindkét esetben a tényleges vizsgálatba vont adatok köre kisebb, mint az elméletileg meghatározott adatoké, illetve minőségük sem tekinthető tökéletesnek. A kutatónak az eredmények interpretációjánál szem előtt kell tartania az adatbázisok ezen torzító hatásait, és bizonyos esetekben mérlegelnie kell több adatbázisból származó adatok együttes használatát. Ez nem minden esetben problémamentes, hiszen az adatok harmonizálása, egymásnak megfeleltetése újabb nehézségeket eredményezhet.

A tudománymetria leggyakrabban bibliográfiai metaadatokra támaszkodik, a jellemzően nagy tömegű (pl. több ezertől akár több millióig terjedő publikációszám) adat vizsgálatát legtöbbször kvantitatív módszerekkel végzik. A vizsgálatokhoz az adatokat általában valamilyen nemzeti vagy nemzetközi ún. citációs adatbázisból érik el a kutatók. A két legismertebb és leginkább a tudománymetriai elemzésekhez használt nemzetközi citációs adatbázis a Scopus és a Web of Science (WoS). A kutatásokhoz felhasznált adatokat rendszerint egy-egy tudományágra, illetve egy néhány éves időintervallumra szűrik. Például Hoekman és munkatársai (2010) az együttműködések és a távolság kapcsolatát vizsgálták a WoS 2000 és 2007 között közel 9000 folyóiratában megjelent társszerzős publikációkon, Luukkonen és munkatársai pedig (1993) a nemzetközi együttműködések a Science Citation Index és a Social Science Citation Index 1981 és 1986-os adatai alapján vizsgálták. Előfordul, hogy egy kutatás során egy vagy két folyóirat valamennyi cikkének feldolgozása történik, pl. Ahlgren és Jarneving (2008) a WoS adatbázisból egyetlen folyóirat, az *Information Retrieval* 2004 és 2006 közötti cikkeit elemezte a bibliográfiai csatolás módszerével. A tudománytérképek létrehozásához pedig gyakorlatilag az egész WoS tartalmát használták (Leydesdorff és Rafols, 2009; Rafols et al., 2010; Leydesdorff et al., 2013).

5.2. A tudománymetriában leggyakrabban használt adatbázisok rövid bemutatása

Ebben a fejezetben a tudománymetriai vizsgálatokban leggyakrabban használt két nemzetközi citációs adatbázist, a Web of Science-et (WoS) és a Scopust mutatom be röviden, ugyanis e tudományterület kutatásai legnagyobbbrészt a WoS, kisebb részt a Scopus és egyéb adatbázisok (pl. Google Scholar, nemzeti adatbázisok) adataira támaszkodnak (Rafols, 2017 STI plenáris előadás).

A WoS a legrégebbi citációs adatbázis, mely az Eugene Garfield által az 1960-as években megalkotott Science Citation Index adataiból nőtte ki magát (Aghaei Chadegani et al., 2013). Az adatbázis 2004-ig, a Scopus megjelenésig, egyeduralkodó volt a piacon (Bar-Ilan, 2008; Vieira és Gomes, 2009; Mongeon és Paul-Hus, 2014). A WoS egy online kutatási platform, amelynek a tulajdonosa előbb a Thomson Reuters, majd 2016. október 3. óta a Clarivate Analytics. Az adatbázis hét citációs adatbázist foglal magában: Science Citation Index Expanded, Social Science Citation Index, Art & Humanities Citation Index, Index Chemicus, Current Chemical Reactions, Book Citation Index és Conference Proceedings Citation Index (wokinfo.com). Az adatbázis tartalma kiterjed a három fő tudományterület, úgymint természettudományi, társadalomtudományi és bölcsészettudományi témájú, a két előbbi terület esetében 1900, a bölcsészettudomány esetében 1975 óta megjelent publikációk citációs adataira. Jelenleg több mint 18 000 magas presztízsű folyóirat publikációit tartalmazza, melyek az egész világról bekerülhetnek az adatbázisba (www.clarivate.com). Igaz, a tartalmazott folyóiratokon belül az angolszász területek felülreprezentáltak, és egészében véve is angol nyelvű folyóiratok szerepelnek többségében az adatbázisban. A WoS Core Collection a három fő tudományterület legrangosabb folyóiratainak a gyűjteménye a WoS-on belül. Az utóbbi években a WoS egy újabb multidiszciplináris citációs indexszel, az Emerging Sources Citation Index-szel (ESCI) bővült, mely mind a három nagy területről tartalmaz folyóiratot. Az ESCI-ben bennfoglalt folyóiratok a későbbiekben a három kiemelt citációs indexbe is bekerülhetnek, a három kiemelt citációs indexből kikerülő folyóiratok pedig az ESCI-be kerülnek (Testa, 2016). A *Tér és Társadalom* folyóirat nemrégiben (2017) az ESCI részévé vált (www.rkk.hu/hu). Az ESCI lehetővé teszi, hogy az értékelési eljárás alatt álló folyóiratok is láthatóvá váljanak a WoS-ban.

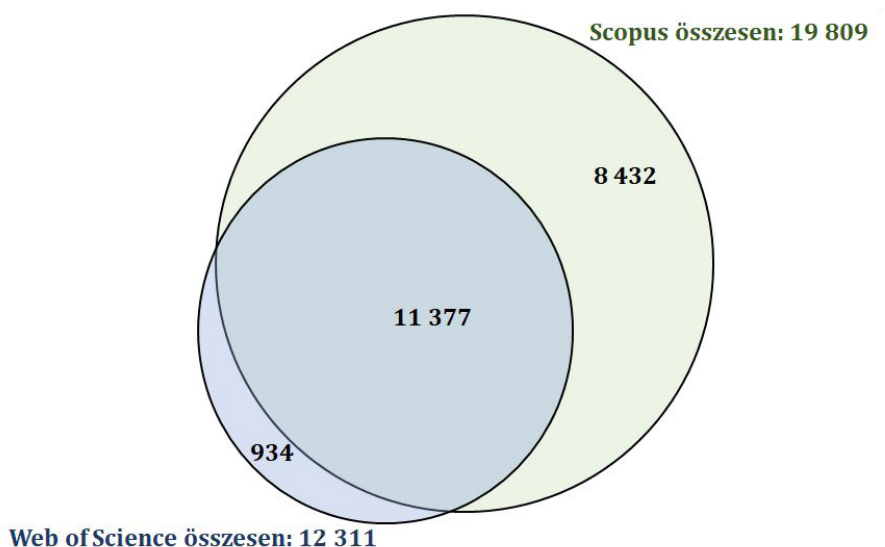
Az adatbázisban szereplő folyóiratok minden egyes publikációjára sok más mellett a következő főbb adatokat találhatjuk meg: cím, szerző(k) neve, szerzői affiliációk, megjelenés éve, folyóirat címe, citációs indexek, kulcsszavak, a cikkek végén található hivatkozott dokumentumok adatai (*cited references*), különböző részletezettségű tudományterületi besorolás (Web of Science Category [WCat], Subject Areas stb.). Citációs adatbázisként egy-egy publikáció esetében közvetlenül lekérdezhető, hogy az adott cikket mely más publikációk idézték, valamint az adott publikáció rekordja tartalmazza a cikk által hivatkozott publikációk adatait. Az adatbázison belül a hivatkozó és hivatkozott publikációk össze vannak egymással kötve.

A WoS-ban elérhető folyóiratokhoz tudományterületi besorolás tartozik, az egyes publikációk tudományterületi besorolása a folyóiratokon keresztül történik, azok besorolásáról „öröklődik” a publikációkra, így a cikk tudományterületi besorolását az azt bennfoglaló folyóirat határozza meg. Különböző

részletezettségű tudományterületi besorolások szerepelnek a WoS-ban, melyek közül a legaggregáltabb szint 5 kategóriát tartalmaz, ezek az élettudományok és orvosbiológia (*Life Sciences & Biomedicine*), a fizikai tudományok (*Physical Sciences*), a műszaki tudományok (*Technology*), a bölcsészettudományok (*Arts & Humanities*), illetve a társadalomtudományok (*Social Sciences*) (apps.webofknowledge.com). A tudományterületi besorolások közül a Web of Science Category (WCat) a legrészletesebb, közel 250 tudományterületet különböztet meg. Ebben egy folyóirathoz több WCat is tartozik.

A másik nagy citációs adatbázis, a Scopus az Elsevier tulajdonában álló lektorált tudományos folyóiratok, könyvek és konferenciakötetek metaadatait és citációs adatait tartalmazza (www.open.ac.uk/blogs/the_orb/?p=2062). Az adatbázisban több mint 22 800 folyóirat és 150 000 könyv adatai találhatóak meg. A Scopus is tartalmazza a publikációkra vonatkozó hivatkozásokat (*cited references*) egészen 1970-ig visszamenőleg (Scopus Content Coverage Guide, 2017, www.elsevier.com), valamint a hivatkozott és hivatkozó publikációk ebben az adatbázisban is össze vannak kapcsolva (Nagy, 2016). Az adatok minden tudományterületet lefednek, úgymint természettudomány, matematika, mérnöki tudományok, egészség- és orvosi tudományok, társadalom- és bölcsészettudományok (www.elsevier.com). A négy nagy tudományterület 27 kisebb tudományterületre (Subject Area Classification) oszlik (service.elsevier.com). Emellett a legrészletesebb szint 334 altudományterületet tartalmaz (ASJC – All Subject Journal Classification Codes). Az adatbázisban egy-egy publikációról többek között a következő adatok érhetők el: szerző(k) nevei, szerzők affiliációja, dokumentum címe, megjelenés éve, forrás címe (pl. folyóirat), dokumentum típusa, hivatkozások száma. A citációk vagy hivatkozások száma csupán a Scopus adatbázisában indexelt tartalmakra vonatkozóan kerül számításra, azaz ha valamilyen Scopusban megtalálható dokumentum idézte az adott dokumentumot (service.elsevier.com). Az adatbázis többek között a kutatói profilokhoz tartozó h-indexek kiszámítását is lehetővé teszi.

A két adatbázis tartalma erősen átfed egymással. Általánosságban elmondható, hogy a Scopus több folyóiratot tartalmaz, míg a WoS a jelentősebb, jobb hatásmutatókkal rendelkező folyóiratokat gyűjti inkább össze. A 12. ábra a WoS és a Scopus tartalmának egymáshoz viszonyított arányát mutatja 2013-as adatokon (Aghaei Chadegani et al., 2013.)



12. ábra: A WoS és a Scopus tartalmának viszonya egymáshoz képest 2013-ban

Forrás: Aghaei Chadegani et al., 2013, 19. o. alapján saját szerkesztés

5.3. A kutatásba vont adatok

Doktori kutatásomban, az 1.2. fejezetben részletezett kutatási kérdések alapján, a különböző távolságtípusok azonosítását, elkülönítését, jellemzőinek bemutatását és egymáshoz való viszonyát vizsgálom tudományos hálózatokon. A kutatás során a tudományos együttműködési hálózatok közül a társszerzői hálózatokat veszem górcső alá, melyekben a kutatói együttműködés publikációk létrehozására irányul. A tudományometriában az alapvető vizsgálati entitás a publikáció. A citációs adatbázisokban az adatrekordok is az egyes publikációkra vonatkoznak. A különböző vizsgálatok azonban különböző aggregációs szinteken történhetnek, mint például a szerzők (Melin és Persson, 1996; Zhao és Strotmann, 2008a, 2008b), a cikkek (Luukkonen et al., 1993; Ahlgren és Jarneving, 2008; Hoekman et al., 2010) vagy a folyóiratok (Vida, 2016; Rahman et al., 2017) szintje. E szintek közül tehát alapszintnek a cikkek szintje tekinthető, a legtöbb elemzés is erre vonatkozik.

Az együttműködési hálózatok vizsgálata is különböző aggregációs szinteken értelmezhető, melyeken a kapcsolatot a hálózat résztvevői között a közös publikációk teremtik meg. A társszerzői hálózatok esetében az alapszint a szerzők és cikkek szintje. Ekkor a csúcsok a szerzők, az él(ek) pedig a közös cikk(ek) a két szerző között. Azonban a szerzői affiliáció révén lehetőség van aggregáltabb formában is vizsgálni az együttműködési hálózatokat. A szerzői affiliáció a cikkben a szerző által megadott intézmény neve és címe. A WoS

szerzői affiliációs adatait tartalmazó mezőben található a szerző intézményének a neve (legtöbbször az osztály vagy tanszék nevével kiegészülve), valamint a település és az ország neve irányítószámmal: olykor az a megye vagy régió is szerepel, ahol az intézmény található. A nemzetközi együttműködések vizsgálata ezen adatokon keresztül történhet (Schubert és Braun, 1990; Braun et al., 1992; Luukkonen et al., 1993). Az együttműködési hálózatok vizsgálata az adatok felhasználásával különböző szinteken történhet (Katz és Martin, 1997): ekkor a hálózat a cikkek és a különböző szerzői affiliációs szintek közötti kapcsolatot mutatja, például a cikkek és a szerzők (Melin és Persson, 1996), a cikkek és az intézmények (Martin et al., 1990; Katz, 1992), a cikkek és a települések (Grossetti et al., 2014) vagy a cikkek és az országok viszonylatában (Moed et al., 1991; Luukkonen et al., 1993). A hálózatok vizsgálata egyre aggregáltabb szintek alapján történhet a szerzőktől az országokig.

Jelen doktori kutatásban az alapszintet, a szerzői szintet választottam. Társ-szerzői hálózatomban két szerző között akkor szerepel él, ha legalább egy közös cikkük volt a vizsgált időszakban. A létrehozott hálózat irányítatlan és súlyozott. Irányítatlan, hiszen a társszerzőség egy-egy cikk esetében kölcsönös kapcsolatot feltételez, és súlyozott, mivel ha két szerzőnek több közös cikke is volt a vizsgált időszakban, akkor közöttük erősebb kapcsolatot feltételezhetünk. A 4.3. fejezetben a konceptuális modellem leírásakor a társadalmi komponenst a társszerzői hálózattal határoztam meg, a kognitív komponenst pedig a szerzők közötti bibliográfiai csatolással (ld. 4.2. fejezet). A szerzői affiliációs adatokat a szerzők közötti fizikai távolság vizsgálatához használtam fel.

A 4.3. fejezet konceptuális modelljében meghatározott három távolságtípust két tudományterületen vizsgáltam meg a WoS adatbázisban 2010 és 2014 között publikált cikkek adatain. Azokat a tudományos publikációkat azonosítottam 2010 és 2014 között, amelyek a választott két tudományterület egyikebe kerültek besorolásra, illetve a szerzők között legalább egy magyarországi affiliációval rendelkező szerepelt.

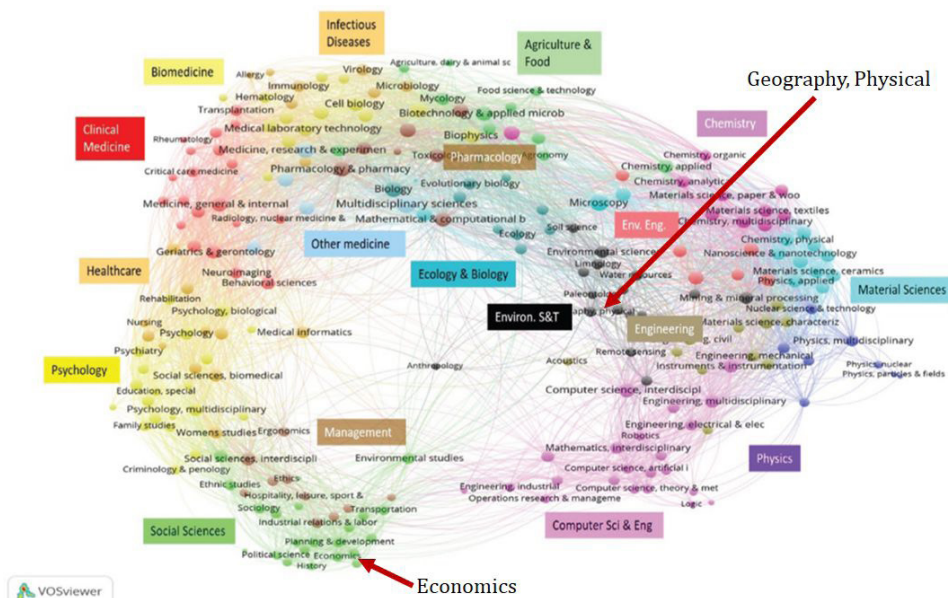
Közismert, hogy a különböző tudományterületek eltérő publikációs jellemzőkkel bírnak (Larivière et al., 2006), például a társadalomtudományok, különösképpen a bölcsészettudományok területén a könyv vagy monográfia jellemző publikációs típus, szemben a természettudományi területen gyakoribb folyóiratcikkkel (Hicks, 1999, 2004). Ezek az eltérő publikációs jellemzők a tudományos együttműködésekben is megmutatkoznak, például a természettudományok területén jellemző a magasabb szerzőszám. Az értekezésben egy társadalomtudományi és egy természettudományi területet vontam be a vizsgálatba, mivel a különböző szakterületeken az együttműködésekben kirajzolódó mintázatokra voltam kíváncsi. A választott két tudományterület

a természetföldrajz (*Geography, Physical*) és a közgazdaságtan (*Economics*) volt. A választásom azért erre a két tudományterületre esett, mert geográfusként mindkettő közel áll hozzám. Társadalomföldrajz helyett azért választottam a közgazdaságtant, mert az sokkal nagyobb számban van jelen a WoS adatbázisban, jobban beazonosítható és mégis viszonylag szoros kapcsolatot mutat a társadalomföldrajzzal. A WCat-ek között külön társadalomföldrajzra vonatkozó kategória nem szerepel. Amikor pedig a *Geography* WCat-re kerestem rá az adatbázisban, akkor a találati listában számos, a *Geography, Physical*-ban is jelen levő természetföldrajzi publikációs rekordot találtam. A *Geography* kategória ugyan tartalmazta a társadalomföldrajzhoz kapcsolódó publikációkat, de a természetföldrajzhoz tartozó publikációk is megtalálhatók voltak benne. A két WCat-ből végül ezért a természetföldrajz mellett egy másik tudományterületet választottam, mely jól elkülönül a természetföldrajztól, viszont a társadalomföldrajzhoz valamilyen szinten közelebb áll. A kiválasztott két tudományág lehetővé tette annak vizsgálatát, hogy egy természet- és egy társadalomtudomány eltérő jellegzetességeket mutat-e a különböző távolságdimenziók mentén.

A 13. ábra egy tudománytérképet mutat be. A szerzők (Carley et al., 2017) 18 nagyobb tudományterületi klasztert különítettek el a tudománytérképen, mely tulajdonképpen a cikkek közötti hivatkozásokra épülő hálózat. A természetföldrajz a (térképen feketével jelölt) környezettudomány klaszterébe került, míg a közgazdaságtan a (zölddel jelölt) társadalomtudomány klaszterében található. A térképen látható, hogy míg a természetföldrajz a többi természettudományi területhez, addig a közgazdaságtan a többi társadalomtudományhoz található közelebb.

Általánosságban elmondható, hogy a tudományometriában a természettudományok vizsgálata jóval gyakoribb, ennek egyik oka, hogy e területek nagyobb arányban találhatók meg a nemzetközi citációs adatbázisokban (Jacso, 2005). A hazai társadalom- és bölcsészettudományok kevésbé vannak jelen pl. a WoS-ban: ennek egyik oka, hogy míg a természettudományok esetében az angol nyelvi közlés kevésbé okoz problémát, addig például a humán tudományok területén a nyelv fontos közvetítő elem, jobban meghatározza a tartalmat (Nagy, 2016).

Egy-egy tudományterület eltérő publikációs és hivatkozási szokásokkal bír, például az orvostudományban jellemzőek a rövid, egy-egy konkrét klinikai kutatásra vonatkozó cikkek, a társadalom- és bölcsészettudományok területén pedig szokványosabb a nagyobb lélegzetű művek publikálása.



13. ábra: A természetföldrajz (Geography, Physical) és a közgazdaságtan (Economics) elhelyezkedése egy 2015-ös tudománytérképen. A térkép a WoS 227 tudományterületi kategóriája alapján 18 klasztert különít el

Forrás: Carley et al., 2017, 87. o.

Ebből következően az előbbi tudományterületen időegység alatt több, míg az utóbbi területen kevesebb publikáció keletkezik. Emiatt nagyon fontos, hogy a különböző tudományometriai vizsgálatok, ezen belül pedig különösen a hatásmérés egy-egy tudományterületen belül vonjon le következtetéseket, illetve valamilyen módon sztenderdizált mutatók révén hasonlítsa össze egymással az eltérő tudományterületeket (Bornmann és Marx, 2013).

A természet- és társadalomtudományok társszerzői hálózatai is gyakran különböznek egymástól, például a természettudományok esetén a társszerzőségek jóval gyakoribbak, mint a társadalomtudományok esetében (Newman, 2001a; Mali et al., 2012). A biokémia vagy a részecskefizika területén megjelenő publikációk gyakran több száz, akár több ezer szerzővel rendelkeznek (Cronin, 2001; Glänzel és Schubert, 2004). Ezeket egy elemzés során érdemes külön kezelni, melyre a tudományterületi bontás alkalmasnak bizonyulhat.

A választott két tudományterületbe tartozó cikkek azonosításához a WCat-eket alkalmaztam. Amikor kizárólag a közgazdaságtanra (*Economics*) és a természetföldrajzra (*Geography, Physical*) vonatkozólag történt a lekérdezés, a cikkek számában jelentős eltérés mutatkozott a közgazdaságtan javára (693, illetve 162). Így végül a tudományterületi lehatárolást nem egy konkrét WCat mentén végeztem, hanem egy WCat-csoportot használtam. Az egymással kapcsolatban levő WCat-ek azonosítását hierarchikus klaszterelemzéssel végeztem a WCat-ek és folyóiratok együtt-előfordulási távolságmátrixán. Először a Journal Citation Report-ból (JCR) legyűjtöttem a WoS-ban található folyóiratokhoz tartozó WCat-besorolásokat. Utána létrehoztam egy együtt-előfordulási mátrixot a közös WCat-ek alapján (Vida, 2016). Az együtt-előfordulási mátrix tulajdonképpen egy hasonlósági mátrix volt, melyben két WCat annál közelebb állt egymáshoz, minél több folyóirat besorolásában szerepeltek együtt. A hasonlósági mátrixhoz a Salton-féle koszinusz hasonlóságot használtam (Salton és McGill, 1983). A mátrix értékei így 0 és 1 közé estek, és a cellaérték minél magasabb volt, annál nagyobb hasonlóság állt fenn a két WCat között. A hasonlósági mátrixból definiáltam a távolságmátrixot, a következő képlettel:

$$D = 1 - S$$

ahol D a távolságmátrix és S a hasonlósági mátrix. Ezt követően hierarchikus klaszterezéssel WCat-csoportokat különítettem el a távolságmátrixon belül. A természetföldrajz esetében a 80%-os, a közgazdaságtan esetében a 90%-os vágási szintet választottam, mivel ezeken a szinteken a választott tudományterülettel egy klaszterbe eső tudományterületek nem kizárólag a természetföldrajzra vagy a közgazdaságtanra korlátozódtak, de túl sok más tudományterület sem szerepelt még a választott területek mellett. A két tudományterület-csoport meghatározása tehát a két vágási szinten a választott tudományterületeket tartalmazó klaszter alapján történt. A választott csoportba tartozó WCat-ek a következők voltak:

Közgazdaságtan:

- Agricultural Economics & Policy (Agrárgazdaságtan, mezőgazdasági politika)
- Business, Finance (Üzleti tudomány, pénzügy)
- Economics (Közgazdaságtan)

Természetföldrajz:

- Geography, Physical (Természetföldrajz)
- Geosciences, Multidisciplinary (Multidiszciplináris földtudományok)
- Imaging Science & Photographic Technology (Képkészítő eljárások, fotó-technológia)
- Remote Sensing (Távérzékelés)
- Engineering, Geological (Mérnökgeológia)

A továbbiakban, amikor a természetföldrajz vagy a közgazdaságtan tudományterületet említem, mindig az egész WCat-csoportot értem alatta.

Összefoglalva, a vizsgálat alapkorpuszába a 2010 és 2014 között megjelent tudományos publikációk közül azok tartoztak, melyeknek legalább az egyik szerzője magyar affiliációval rendelkezett, illetve a két választott tudományterülethez került besorolásra. A szűrés eredményeként a közgazdaságtan területéről 722 cikk, míg a természetföldrajz területéről 653 cikk WoS rekordját töltöttem le az adatbázisból.

Egy WoS rekord egy több mint 60 mezőt, az egyes cikkek legfontosabb ismertetőjegyeit tartalmazó adathalmaz. A 2. táblázat egy rövid mintát mutat ebből. Vizsgálatomhoz a következő mezők voltak szükségesek: cikk WoS azonosítója (UT), szerző neve (AU), szerző affiliációja (C1), cikk hivatkozott cikkeinek az adatai, azaz a cikk végén szereplő felhasznált irodalom tételei (CR). Ezen mezők mellett kiegészítő információként használtam még a cikk címét (TI), folyóirat nevét (SO), cikk kulcsszavait (DE), illetve az adattisztításhoz az (AF) mezőt is, mely a szerzők teljes nevét tartalmazza. Az adatok előkészítését az 5.4.2., konkrét felhasználásukat a 6. fejezet tartalmazza.

UT	AU	AF	C1	CR	TI	SO	DE
WOS:000348908000001	Mero, K	Mero, Katalin	IBS, Dept Finance & Accounting, Budapest, Hungary	Baker A, 2013, NEW POLIT ECON, V18, P112, DOI 10.1080/13563467.2012.662952; Borio C., 2003, BANK INT SETTLEMENTS	MACROPRUDENTIAL WARNINGS IN THE EURO ZONE AND HUNGARY	ACTA OECONOMICA	macroprudential warnings; crisis; financial supervision; financial institutions; Hungary
WOS:000348908000002	Virag, M; Nyitrai, T	Virag, Miklos; Nyitrai, Tamas	[Virag, Miklos] Corvinus Univ Budapest, Dept Enterprise Finances, Budapest, Hungary; [Nyitrai, Tamas] Corvinus Univ Budapest, Budapest, Hungary	Ahn BS, 2000, EXPERT SYST APPL, V18, P65, DOI 10.1016/S0957-4174(99)00053-6; ALTMAN EI, 1968, J FINANC, V23, P589, DOI 10.2307/2978933	IS THERE A TRADE-OFF BETWEEN THE PREDICTIVE POWER AND THE INTERPRETABILITY OF BANKRUPTCY MODELS? THE CASE OF THE FIRST HUNGARIAN BANKRUPTCY PREDICTION MODEL	ACTA OECONOMICA	bankruptcy prediction; data preparation; outliers; discretisation; support vector machines (SVM); rough set theory (RST)
WOS:000347758800001	Endresz, M; Harasztosi, P	Endresz, Marianna; Harasztosi, Peter	[Endresz, Marianna; Harasztosi, Peter] Magyar Nemzeti Bank, H-1054 Budapest, Hungary	Aghion P, 2001, EUR ECON REV, V45, P1121, DOI 10.1016/S0014-2921(00)00100-8; Basso H.S., 2007, 748 ECB WP; Benavente Jose M., 2003, EMERG MARK REV, V4, P397, DOI 10.1016/S1566-0141(03)00062-1	Corporate foreign currency borrowing and investment: The case of Hungary	EMERGING MARKETS REVIEW	Liability dollarization; Currency mismatch; Investment; Balance sheet effects; Liquidity constraints

2. táblázat: WoS rekord felhasznált mezői (példa)

A vizsgálatba csak azokat a cikkeket vontam be, melyekben az idézett hivatkozások tartalmazó CR mező nem volt üres, mivel a felhasznált irodalmi tételek közötti hasonlóság definiálja a teljes kognitív távolságot (lásd 6.2.1. alfejezet). A CR mező a természetföldrajz esetében a cikkek 99%-ában, míg a közgazdaságtan esetében a cikkek 65%-ában szerepelt a WoS rekordban. Sajnos az ebből adódó esetleges torzítást el kell fogadnunk a két tudományterület eredményeinek összehasonlításakor. Az egy tudományterületen belüli három távolságmátrix összehasonlítását ez nem érinti.

Ennél is nagyobb eltérést mutatott a szerzők száma a két tudományterület között. Míg a cikkek száma közel azonos volt a két tudományterületen, a szerzők számában már nagyságrendi eltérést tapasztaltam (3. táblázat). A társszerzős cikkek aránya 61%, illetve 89%, míg az egy cikkre eső átlagos szerzőszám 3, illetve 6 volt a közgazdaságtani és a természetföldrajzi cikkek esetében. A természettudományok területén a társszerzős cikkek aránya általában rendre magasabb (Glänzel és Schubert, 2004), a jelen minta is ezt támasztotta alá.

	Közgazdaságtan	Természetföldrajz
Cikkek száma	722	653
Cited References mezővel rendelkező cikkek száma	470	644
Szerzők száma	704	2294

3. táblázat: Az adatbázis mérete – magyar affiliációval is rendelkező cikkek 2010 és 2014 között a két tudományterületen

E jelenség miatt a társadalmi komponens várhatóan meghatározóbb lesz a teljes kognitív távolságban a természetföldrajz területén. Mivel a tiszta kognitív távolság a társszerzőségekből adódó torzító tényező nélkül mutatja a szerzők közötti kognitív kapcsolatokat, ott ez a hatás valószínűleg nem lesz kimutatható.

5.4. Az adatok felhasználásának korlátai

Vizsgálatomhoz a WoS adatbázis bizonyult a legmegfelelőbbnek, annak ellenére, hogy mint minden adatbázisnak, ennek is vannak hibái. Előnye, hogy ez a legrégebbi nemzetközi citációs adatbázis (Bar-Ilan, 2008; Vieira és Gomes 2009; Aghaei Chadegani et al., 2013; Mongeon és Paul-Hus, 2014), illetve minden szükséges adatot megbízható, jól elemezhető formában érhettem el a doktori kutatásomhoz. Azonban több probléma is felmerült az adatokkal kapcsolatban. Ezeket két nagy csoportra bonthatjuk.

A nehézségek egyik csoportja magából az adatbázisból következik. Az elérhető adatok miatt a vizsgálat nem a létező összes 2010 és 2014 között, legalább egy magyar szerzővel rendelkező, a természetföldrajz vagy közgazdaságtan területén megjelent tudományos publikációra terjedhetett ki, csupán azokra a cikkekre, melyek a WoS-ban megtalálhatók voltak. Ezt az eredmények értelmezésekor fontos figyelembe venni. A szakirodalomból ismertek a WoS tartalmából következő ún. torzító hatások, mint például az angol nyelv túlsúlya, a földrajzi lefedettségben található aránytalanságok, a folyóirat-központúság, valamint a természettudományok relatív felülreprezentáltsága (Moed et al., 1991; Leeuwen et al., 2001; Jacso, 2005; Mongeon és Paul-Hus, 2014).

A másik problémakör a letöltött adatok feldolgozásakor válik nyilvánvalóvá, mint például bizonyos adatok hiánya, az adatrekordok hibája (pl. elírás, a hivatkozás információjának nem egységes sorrendben közlése stb.). Ezeket a problémákat az elemzés megkezdése előtt kell kezelni. A következő két alfejezetben ezeket a problémákat és megoldásaikat ismertetem.

5.4.1. A WoS adatbázisból adódó torzítások

Ahhoz, hogy a WoS adatbázis tartalmából eredő torzító hatásokat megértsük, fontos, hogy röviden megismerjük az adatbázisba bekerülő folyóiratok kiválasztási mechanizmusát. A WoS-ben szereplő folyóiratok gondos válogatás eredményeképpen kerültek be az adatbázisba. A legjelentősebb folyóiratok a WoS Core Collection részét képezik. Ebben a három fő citációs index folyóiratai szerepelnek összegyűjtve, mint Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) és Arts & Humanities Citation Index (AHCI). A három citációs index folyóiratai lefedik a természet-, a társadalom- és a bölcsészettudományokat: ezek tekinthetők a legjelentősebb folyóiratoknak a világon. Az SCIE és az SSCI folyóiratokra a Journal Citation Reportban minden évben közlik a folyóirat impakt faktorát.

A kiválasztásnál ugyanazon kritériumoknak kell megfelelnie egy folyóiratnak az ESCI-be (a WoS legújabb citációs indexébe) való bekerüléshez, mint az SCIE, a SSCI és az AHCI esetén, egyetlen különbséggel: a citációs impakt alacsonyabb értékű is lehet. Ezek a kritériumok a következők: a folyóirat legyen a peer-review alatt, kövesse az etikus publikálási gyakorlatokat, feleljen meg a WoS technikai követelményeinek, a bibliográfiai információk legyenek angolul elérhetők és a WoS kutatói felhasználói által legyen ajánlva a folyóirat (Testa, 2016).

Eugene Garfield, aki a citációs indexek atyának tekinthető (Nagy, 2016) hozta létre a mai SCIE elődjét, a Science Citation Indexet. Felismerte, hogy létezik egy ún. „központi irodalom” (*core literature*), mely a legjelentősebb folyóirato-

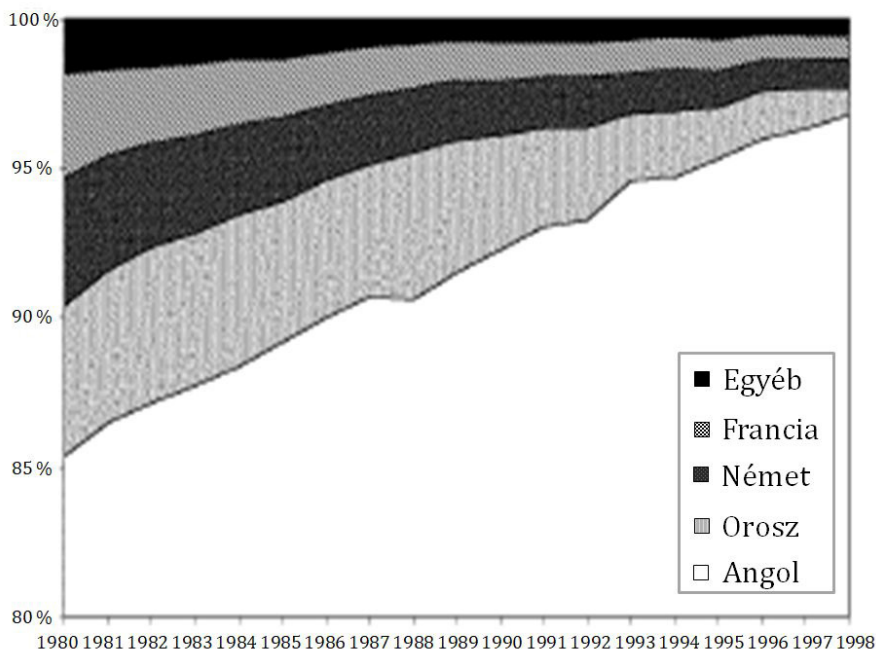
kat foglalja magába. A tudományos tudás legnagyobb része néhány folyóiratban koncentráltan jelenik meg, ez az egyes tudományterületekre külön-külön is igaz. Ezt a jelenséget Garfield koncentrációs törvényének (*Garfield's law of concentration*) is nevezik (Garfield, 1971). Ezt a központi magot jeleníti meg a WoS-ban a Core Collection, mely 2014-ben közel 12 000 folyóiratból állt, ezek közül a cikkek 80%-a 4470 folyóiratban jelent meg, de a hivatkozások 85%-a is ezekben a folyóiratokban megjelenő cikkekre történt (Testa, 2016). Fontos továbbá még, hogy ez a mag nem állandó, folyamatosan változik egyes kutatási kérdések relevánsabbá válásával, mások jelentőségének csökkenésével.

Mindezekből arra következtethetünk, hogy igaz ugyan, hogy a WoS a tudományos publikációknak csupán a minőségi elitjét tartalmazza, de Garfield koncentrációs törvénye is alátámasztja, hogy az adatbázis a tudományos tudás leginkább hivatkozott, és ezáltal a tudomány éppen aktuális témáival foglalkozó folyóiratait foglalja magába.

A kiválasztási procedura azonban akármennyire is körültekintő és széleskörű próbál lenni, eredményez torzító hatásokat. A legjelentősebb torzító hatás a nyelv, azon belül is az angol nyelv túlsúlya (Moed et al., 1991). Bár a tudomány nyelvének az angol tekinthető, vannak olyan területek, ahol az attól eltérő anyanyelvi közlésben megjelent munkák nagyon jelentős eredményeket tartalmaznak, de a nyelvi akadályok miatt a világ tudományos vérkeringésétől elzárva maradnak. A bölcsészettudományok területén például egy-egy irodalmi mű elemzése a mű keletkezési nyelvén sokkal nagyobb érdeklődésre tarthat számot. Az alternatív gyógyászatról a legtöbb cikk kínaiul jelenik meg, ahol a tudományterületnek nagy tradíciója van, ez azonban a WoS-ban láthatatlan marad (Fan 2015), mivel a folyóirat adatbázisba történő bekerüléséhez legalább a bibliográfiai adatoknak angolul kell szerepelniük.

Fan (2015) az alternatív gyógyászattal foglalkozó folyóiratok jelenlétét vizsgálta a WoS-ban 2003 és 2013 között. Azt találta, hogy az alternatív gyógyászattal foglalkozó folyóiratok száma valamelyest nőtt, viszont a WoS-ban újonnan elérhető folyóiratok nagyrészt a nyugati világban kerültek kiadásra (főképp az Egyesült Államokban és az Egyesült Királyságban). Ez egyúttal példa a földrajzi lefedettségben jelenlevő aránytalanságokra is. Leeuwen és munkatársai (2001) szintén felhívták a figyelmet a WoS-ban található nyelvi torzításra: az elemzésükből kirajzolódó kép (14. ábra) a Science Citation Index (SCI) nyelvi lefedettségét mutatja 1980 és 1998 között. Látható, hogy az angol nyelvű cikkek aránya folyamatosan nőtt az adott időszakban. Leeuwen és munkatársai (2001) azt is kimutatták, hogy az angol nyelven írt cikkekre többen hivatkoznak. Összehasonlították néhány ország publikációs teljesítményét, és amennyiben például Svájcnak vagy Németországnak az SCI-ben látható teljes vagy csak az angol nyelvű publikációs teljesítményét arányosít-

va vizsgálták, a csak angol nyelvű publikációk figyelembevételkor az országok közötti rangsorban jobb helyezést értek el ezek az országok. A csak angol nyelvű publikációk esetében Németország az Egyesült Királysággal azonos szinten helyezkedett el a rangsorban, míg Svájc az első helyezett lett az USA-t is megelőzve.



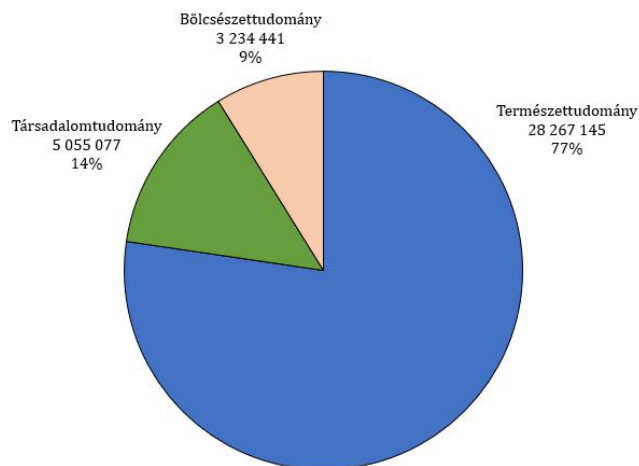
14. ábra: Nyelvi lefedettség az SCI-ben 1980 és 1998 között

Forrás: Leeuwen et. al., 2001, 344. o. alapján saját szerkesztés

Az előzőekből kitűnik, hogy a nyelvi torzítás két szinten is jelentkezik, egyrészt magában az adatbázisban történő láthatóságban, másrészt a cikkek hatásmutatóiban (Leeuwen et al. 2001; Cameron, 2005; Aghaei Chadegani et al., 2013; Liang et al. 2013; Mongeon és Paul-Hus, 2014). Cameron (2005) azt is hangsúlyozta az impakt faktor többoldalú kritikájával foglalkozó tanulmányában, hogy az angol nyelvű cikkek magasabb hatásmutatói miatt sok nem angol anyanyelvű kutató olyan angol nyelvű, magas presztízsű folyóiratban is publikál, amely folyóirat akár nem is elérhető a kutató országában. Általánosságban is elmondható, hogy az angol nyelvű cikkeket többen idézik, így ezek a cikkek jobb hatásmutatókkal rendelkeznek, ennek háttérében az ún. nyelvi határ (*language barrier*) áll. A tudomány nyelve az angol, és a kutatók többsége ezen a nyelven olvas, így ezeket a munkákat is idézi (Liang et al., 2013).

A WoS törekszik arra, hogy a bekerülő folyóiratok földrajzi lefedettsége széleskörű legyen. Ugyanakkor az angol nyelvű publikációk többsége, valamint azok magasabb hatásmutatói a földrajzi lefedettségben is okoznak torzítást, nevezetesen érzékelhető az angolszász kultúra túlsúlya. A WoS és a Scopus adatbázisban található folyóiratok legtöbbje az Egyesült Királysághoz és Hollandiához kötődik, ennek háttérében az áll, hogy a legnagyobb kiadók is ezekben az országokban találhatók (Mongeon és Paul-Hus, 2014).

Egy másik fontos torzító hatás a természettudományi cikkek felülreprezentáltsága (15. ábra) a WoS-ban (Jacso, 2005; Mongeon és Paul-Hus, 2014). Ez a jelenség egyébként a Scopus adatbázisban szintén megfigyelhető. Ezzel összefüggésben még megállapítható, hogy az adatbázis tartalmát legnagyobb-részt folyóiratcikkek alkotják. Ez egyrészt a természettudományi területek felülreprezentáltságából is következik, hiszen e tudományterületek elsődleges közlési csatornáit a folyóiratcikkek. Vannak azonban olyan tudományterületek, ahol a könyvek és monográfiák aránya sokkal nagyobb, ilyenek például a bölcsészettudomány területei. Az adatbázisban szerepelő Book Citation Index e tartalmak láthatóvá tételét segíti elő (Sivertsen, 2009; Nagy, 2016).



15. ábra: A három fő tudományterületi ág rekordjainak megoszlása a WoS-ban

Forrás: Jacso, 2005, 1540. o. alapján saját szerkesztés

Az WoS-ban található ismertett torzító hatásokat az eredmények interpretációjánál figyelembe kell venni. Jelen kutatás szempontjából a torzító hatások közül a nyelv és a földrajzi lefedettség az, ami jelentősebben jelentkezhet. Ez a gyakorlatban azt jelenti, hogy a vizsgálatba vont folyóiratcikkek száma elmarad a legalább egy magyar szerzővel, a választott két tudományterület valamelyikén, 2010 és 2014 között megjelent folyóiratcikkek tényleges szá-

mától, és csak azokat a cikkeket tartalmazza, amelyek a WoS-ban megtalálhatók. Gondolhatnánk, hogy az MTMT mint hazai citációs adatbázis a legalább egy magyar szerzővel rendelkező cikkeket megbízhatóbban, illetve nagyobb arányban tartalmazza, azonban ugyanezen sokaságra nem tudunk közvetlenül lekérdezést eszközölni az MTMT-ből. Ennek oka a cikkekre vonatkozó tudományterületi besorolás hiánya. Az MTMT-ben történő regisztrációkor maga a szerző adhatja meg magáról, hogy mely tudományterülethez tartozónak vallja magát, de a cikkekhez rendelt szakterületi besorolás hiányzik, bár a jövőben tervezik ezen adatok bevitelét is. A másik probléma a szerzői affiliációk nem megfelelő közlése: a szerzői affiliációk sem cikkenként vannak rögzítve, hanem az MTMT-ben regisztrált szerzők esetén csupán a szerzők pályájának egyes periódusaihoz kerülnek hozzárendelésre az intézményi adatok. Az MTMT-ben nem regisztrált társszerzők affiliációi nem elérhetők az adatbázisból. A hivatkozásokat az egyes cikkekhez az MTMT nem tartalmazza, illetve az egyes cikkekre történt hivatkozások adatai is a szerzők révén kerülnek felvitelre. Így a jelen kutatáshoz szükséges adatok közül csupán a társszerzők neveit találhatjuk meg az MTMT-ben. Belátható ezek alapján, hogy bár az elérhető cikkek köre, különösen a közgazdaságtan területén, magasabb lehet az MTMT adatbázisban, azonban sajnos jelen vizsgálathoz az MTMT nem tartalmazza a szükséges metaadatokat.

Doktori vizsgálatom eredményei így a teljes sokaság helyett csupán a WoS-ban található cikkekre vonatkoznak. Ettől függetlenül úgy gondolom, hogy a viszonylag magas elemszám lehetővé teszi az elméleti koncepció alapján a kirajzolódó mintázatok feltárását és a vizsgálati kérdésekre történő válaszok megtalálását.

5.4.2. Az adatok előkészítése a vizsgálat lefolytatásához

Bár a WoS-ból származó adatok jól strukturált formában érhetők el, a feldolgozás során szükséges az esetleges adathiányokat kiszűrni, valamint a jellemzően szöveges adatokat (pl. név) megtisztítani az adatok vizsgálatba történő bevonása előtt. A 5.3. fejezet 2. *táblázata* tartalmazza a letöltött WoS rekordokból a vizsgálathoz szükséges mezőket. Az adatok előfeldolgozásához gépi és kézi eszközöket használtam. A gépi eszközök alatt a R szoftverrel, scripttel történő adatfeldolgozást értem, míg a kézi eszköz a gépi feldolgozást követő ellenőrzést és a scripttel nem megoldható adattisztítást jelentette. A vizsgálatok elvégzéséhez ezt követően szintén R scriptet használtam. Elsősorban három mezőt kellett az alkalmazott módszerekhez előkészíteni: az AU mezőt, mely a szerzők nevét, a CR mezőt, mely a hivatkozott irodalmat és a C1 mezőt, mely a szerzői affiliációkat tartalmazta mindegyik cikkre vonatkozóan.

A társszerzői hálózat létrehozásához a szerzők nevére és a cikkazonosítókat tartalmazó cellára volt szükségem. Ezek közül a szerzők neveit tartalmazó mező tisztítását kellett elvégezni. Az AU mező az adott cikk szerzőinek nevét tartalmazza egymástól pontosvesszővel elválasztott formában (2. táblázat). A neveknél a vezetéknev van megadva, majd vessző és a keresztnév első betűje következik. Több keresztnévet is tartalmazó szerzőknél az összes keresztnév első betűje van feltüntetve egybeírt formában csupa nagybetűvel. Ahhoz, hogy a társszerzői hálózatot meghatározhassam, szükség volt egy két oszlopot tartalmazó tábla kialakítására, amelynél az egyik oszlop a cikkazonosítókat, a másik pedig a szerzők nevét tartalmazta, de itt egy mezőben már csak egy szerző szerepelhetett. Ez azt jelenti, hogy a társszerzős cikkek esetében a pontosvessző mentén „daraboltam” az adatbázisomat. Tehát több szerző esetén minden szerző külön sorba került: mindegyik sor az adott cikkazonosítójával lett ellátva. Ezután szükség volt a nevek tisztítására, ugyanis előfordulhat, főképp szöveges információt tartalmazó celláknál, hogy egy kis különbség a karakterek között a gép számára két külön személyt jelöl, pedig valójában ugyanazon személyről van szó. Ez főképp az esetleges elírásokból, többféle névváltozatból adódik. Az elírások oka lehet például, hogy a magyar nevek olyan karaktereket is tartalmaznak, amiket a WoS adatbázis egy plusz karakterrel helyettesít, pl. ilyenek a hosszú magánhangzók közül az *ő* vagy *ű*. A névváltozatok kiszűrése az, ami miatt igazán szükség van a kézi tisztításra. Ugyanazon személy több névváltozattal történő megjelenése mögött több ok is húzódhat. A szerző maga például nem azonos módon tünteti fel minden cikkben a nevét és emiatt az egyik cikkben csak az első keresztnéve, egy másik cikkben több keresztnéve is szerepel. Női kutatóknál gyakran a házasságot követően férjezett névre változik a vezetéknev. Ezek a problémák leggyakrabban a több vezeték-, illetve keresztnévvel rendelkező szerzőket érintik. Előfordulhat ritkán az is hogy a vezeték- és keresztnév felcserélve kerül be az adatbázisba.

A nevek tisztítása során fontos, hogy minden szerző egy névformával rendelkezzen. Az előzőekben felsorolt példák arra az esetre vonatkoztak, amikor egy valós személy az adatbázisban több néven is szerepelt, ilyenkor ha nem történne korrekció, ezek külön-külön szerzőkként kerülnének azonosításra.

A másik problémakör a gyakori nevek előfordulása. Ekkor egy név valójában több különböző személyt takar. Ezek azonosítása még nehezebb feladat, gyakran nem is lehetséges 100%-os biztonsággal megállapítani, hogy az adott név hány szerzőre vonatkozik. Az AU mező neveinek tisztítása során az adatbázisban található egyéb mezőket is segítségül hívtam. Mint korábban említettem, az AU mező a keresztnévnek csupán az első betűjét tartalmazza, így például Szabó Péter és Szabó Pál ugyanabban a formában (Szabó, P) kerül megjelenítésre. A magyar nyelvben kettős betűk is vannak, ebben az esetben

is a keresztnévnek csak az első karaktere kerül megjelenítésre: így például Szabó Zsófia, Szabó Zsolt és Szabó Zoltán egyformán „Szabó, Z” formában szerepelnek az AU mezőben.

A példákon keresztül látható, hogy egy névalak valójában több személyt jelölhet. Az eredmények szempontjából nem mindegy, hogy egy vagy három szerző adatait elemezzük. Az azonosításhoz a teljes nevet tartalmazó AF mezőre is szükségem volt, ebben a keresztnév is teljes formájában szerepel. Emellett a névtisztítás során használtam még a „C1” szerzői affiliációkat tartalmazó mezőt is. Mivel a vizsgálat egy viszonylag rövid, öt éves periódusra terjedt ki, nagy valószínűséggel egy szerző egy vagy néhány intézményhez volt köthető. A C1 a szerzői affiliációkat szerzőnként mutatja (2. táblázat), azaz szögletes zárójelben látható a szerző teljes neve a keresztnév első betűjével együtt, majd ezt követően az intézmény és annak címe. A nevek tisztításához további segítséget a WoS adatbázis lekérdező felülete adott. Itt lehetőség van egy-egy szerző neve alapján keresni, oly módon, hogy beírjuk a szerző nevét, az adatbázis pedig megjeleníti az összes ahhoz hasonló szerzői névváltozatot, majd lehetőség van a tudományterület, az affiliáció és a szerzőkhöz tartozó cikkek alapján az azonosításra. A kérdéses esetekben e módszerek együttes alkalmazásával végeztem el a nevek tisztítását.

A bibliográfiai csatolás elvégzéséhez szükséges volt a CR mező előkészítése. Ez cikkenként tartalmazza az adott cikk végén szereplő irodalmi tételeket, pontosvesszővel elválasztva. Nem minden cikk esetében volt azonban meg ez az információ. A vizsgálatba csak azokat a cikkeket vontam be, amelyeknél a CR mező nem volt üres. Ez arányaiban a természetföldrajznál volt gyakoribb. Először itt is a pontosvessző mentén „feldaraboltam” a táblát, így kaptam egy másik táblát, amely a cikkazonosítókat és a hivatkozásokat tartalmazta. Egy cikkhez több referencia is tartozik, és a darabolást követően a cikkazonosító annyiszor került külön sorba, ahány hivatkozás szerepelt a cikk végén. A hivatkozások a következő információkat tartalmazzák: szerző neve (teljes vezetéknev és a keresztnév első betűje), megjelenés éve, megjelenés helye (folyóirat neve), kötet, oldalszám, DOI azonosító. Gyakran azonban az itt felsorolt tételek közül nincs meg az összes információ. A legtöbbször a szerző, megjelenés éve és a megjelenés helye került csupán megadásra. Olyan is előfordult, hogy csak néhány karaktert tartalmazott a hivatkozás, ezeket eltávolítottam a létrehozott táblából.

A bibliográfiai csatolás során a cikkek közötti hasonlóságot a felhasznált irodalmi tételek hasonlósága alapján határozzuk meg. Mivel a jelen kutatásban nem a cikkek, hanem a szerzők szintjét elemzem, szerzői bibliográfiai csatolást végzek, a szerzőket és cikkazonosítókat, valamint a cikkazonosítókat és hivatkozásokat tartalmazó táblákat összekapcsoltam a cikkazonosítók segít-

ségével. Így egy olyan táblát kaptam, mely a szerzőket és a hivatkozásokat tartalmazta. Ezt követően a szerzők és a hivatkozások között hajtottam végre a vizsgálatot.

A vizsgálat elvégzése előtt lehetőség van szűrők alkalmazására. Boyack és Klavans (2010) például csak azokat a hivatkozásokat vették figyelembe, amelyeket legalább négyszer hivatkoztak, illetve csak a legalább 5 éve vagy annál régebben hivatkozott cikkekre szűrték az adatokat. Doktori vizsgálatomban nem alkalmaztam ilyen szűrést, mivel míg Boyack és Klavans munkájában több millió hivatkozás szerepelt, addig nálam a hivatkozások száma csupán tízezres nagyságrendű. Az algoritmus gyorsabb futtatása végett eleve csak a legalább kétszer előforduló hivatkozásokat vettem figyelembe, hiszen csak ez esetben volt lehetséges, hogy két cikk is ugyanazt a cikket hivatkozta.

A földrajzi távolság vizsgálatba történő bevonásához szükséges volt a szerzői affiliációkat tartalmazó C1 mező feldolgozása is. Ez tartalmazza a szerzők neveit és a szerzőkhöz tartozó affiliációkat, mint például a szerző intézményének, néhol a tanszékének vagy osztályának, településének, régiójának vagy megyéjének és országának a nevét. A 16. ábra két példát tartalmaz a WoS szerzői affiliációkat tartalmazó mezői közül. A földrajzi távolság vizsgálatához minden egyes szerző affiliációs mezőjéből szükségem volt az intézmény, a település és az ország nevére. A C1 mező cikkenként tartalmazza az információt, így a többszerzős cikkeknél több affiliációt is tartalmaz, illetve az is látható, ha egy cikk több szerzője is ugyanazon intézményhez tartozik. A kívánt cél eléréséhez nagyfokú gépi és kézi adatfeldolgozásra volt szükség. Először a piros vonalak mentén kellett a C1 mezőt „feldarabolni”: ez mutatta meg, hogy egy cikkhez hány különböző intézmény tartozott. A 16. ábra mindkét példájában hét intézmény működött együtt a cikk megírásakor. A gyakorlatban a piros vonalak mentén történő daraboláshoz nem lehetett alkalmazni a pontosvesszőt, mivel akkor egyszerre a nevek és az affiliációk is fel lettek volna vágva.

A piros vonalak helyének a megtalálása a „pontosvessző szóköz nyitó szögletes zárójel” („; [”) mintázat mentén történt. Ezt követően, mielőtt az egyes szerzőkre tovább bontottam volna az adatokat, mindegyik affiliációs sorból kinyertem az intézmény, a település és az ország nevét. A 16. ábrán láthatjuk, hogy az egyes elemeket, mint pl. intézménynév, településnév, ország neve, vessző választja el egymástól. Bizonyos szabályszerűségek is megfigyelhetők, úgymint az első tag mindig az intézmény neve, az utolsó tag pedig az ország neve. Ezek kinyerése volt ily módon a legegyszerűbb. Egyedül az Egyesült Államok esetében volt szükség külön eljárásra, ott ugyanis az ország neve előtt állt az irányítószám. Mivel ezt mindig az USA előtti karaktersorozat tartalmazta, így ezeket viszonylag egyszerűen el lehetett távolítani, automatizált megoldásokkal.

C1
[Markovic, Slobodan B.; Stevens, Thomas; Tomic, Nemanja; Petic, Nikola; Jovanovic, Mladjen; Vasiljevic, Djordjije A.; Gavrilov, Milivoj B.; Obreht, Igor] Univ Novi Sad, Fac Sci, Lab Palaeoenvironm Reconstruct, Novi Sad 321000, Serbia; [Korac, Miomir; Mrdic, Nemanja] Serbian Acad Arts & Sci, Inst Archaeol, Viminacium Archaeol Team, Belgrade 11000, Serbia; [Buylaert, Jan-Pieter; Thiel, Christine] Aarhus Univ, Dept Geosci, Nord Lab Luminescence Dating, DK-4000 Roskilde, Denmark; [Buylaert, Jan-Pieter; Thiel, Christine] Tech Univ Denmark, Ctr Nucl Technol, DK-4000 Roskilde, Denmark; [McLaren, Sue J.] Univ Leicester, Dept Geog, Leicester Quaternary Palaeoenvironm Res Grp, Leicester LE1 7RH, Leics, England; [Stevens, Thomas] Univ London, Dept Geog, Ctr Quaternary Res, Egham TW20 0EX, Surrey, England; [Suemegi, Pal] Univ Szeged, Dept Geol & Palaeontol, H-6722 Szeged, Hungary
[Jull, A. J. Timothy; Lange, Todd E.; Burr, George S.] Univ Arizona, NSF Arizona AMS Lab, Tucson, AZ 85724 USA; [Jull, A. J. Timothy] Univ Arizona, Dept Geosci, Tucson, AZ 85721 USA; [Jull, A. J. Timothy] Inst Nucl Res, H-4001 Debrecen, Hungary; [Panyushkina, Irina P.; Clark, Kelley J.; Salzer, Matthew W.; Leavitt, Steven W.] Univ Arizona, Tree Ring Res Lab, Tucson, AZ 85721 USA; [Kukarskih, Vladimir V.] URAN Inst Plant & Anim Ecol, Lab Dendrochoronol, Ekaterinburg, Russia; [Myglan, Vladimir S.] Siberian Fed Univ, Dept Humanities, Krasnoyarsk, Russia; [Burr, George S.] Natl Taiwan Univ, Dept Geosci, Taipei 10764, Taiwan

16. ábra: Példa a WoS szerzői affiliációkat tartalmazó C1 mező tartalmára

Megjegyzés: A piros vonalak az affiliációkat, a kék vonalak az egy affiliációhoz tartozó szerzőket választják el egymástól.

A településnevek kinyerése bizonyult a legnehezebb feladatnak. A 16. ábra példáján az is látható, hogy a vesszővel elválasztott tagok száma nem azonos, van, ahol szerepel például az osztály vagy tanszék, máshol a megye is szerepel. A települések azonosításában az segített, hogy általában az utolsó előtti tagban kerültek megadásra, de bizonyos országok esetében (mint például az Egyesült Királyság, Kanada, Argentína, Ausztrália és Brazília) az utolsó előtti tag a megyét tartalmazta, a településnév pedig ezt megelőzte. Emiatt külön kellett bontanom a kapott országok körét aszerint, hogy az utolsó előtti vagy a hátulról harmadik tag tartalmazta-e a településnevet. Az Egyesült Királyság esetén a kézi ellenőrzés során találtam olyan sorokat, ahol a megye nem került megadásra, és emiatt az utolsó előtti tag tartalmazta a településnevet. Ezeket az eseteket kézzel korrigáltam.

A kinyert településnevek elől ezt követően el kellett távolítani az irányítószámokat. A számok eltávolítása automatikusan megtehető volt, hisz a településnevekben számok nem szerepeltek, így egy egyszerű cserével megoldható volt a számok kivágása. Sok ország esetében betű is szerepelt az irányítószámokban. Azoknál az országoknál, melyeknél a betűket egy kötőjel követte

(pl. Magyarország H-, Dánia DK-, Németország D- stb.), a kötőjelet megelőző karakterek eltávolításával a felesleges irányítószámot tartalmazó karakterek eltávolítása is megoldható volt.

Itt a legnehezebb feladatot az Egyesült Királyság irányítószámainak eltávolítása jelentette, mivel a betűk között jellemzően szóköz is szerepelt. Végül ezeket kézzel távolítottam el. A településnevek esetében szükség volt a gépi adat-előkészítés után kézzel is tisztítani. Előfordult néhány esetben a településnév elírása, illetve az ékezetes karakterek esetén valamilyen plusz karakter megjelenése. Ez főképp a magyar ékezetes településneveknél volt jellemző, ahol *ő* vagy *ű* szerepelt volna, ahogy ezt már korábban a nevek tisztításánál említettem. Olyan is előfordult, hogy a településnévnél is az intézménynév szerepelt, vagy a település egy részének neve (pl. Miskolc helyett Miskolc-Egyetemváros).

Mivel szerzőnként szerettem volna kinyerni az affiliációkra vonatkozó adatokat, így az affiliációból az intézménynév, településnév és ország nevének kinyerését követően a 16. ábra alapján a kék vonalak mentén is „darabolnom” kellett a táblát. A kék vonalak mentén történő darabolás történhetett a pontosvesszők mentén. Itt egy problémával szembesültem: az egyszerűs cikket külön kellett válogatnom, mivel ezekre már megvolt a kívánt affiliációra vonatkozó információ. A többszerzős cikkeket a pontosvessző mentén daraboltam, ekkor mindegyik szerző új sorba került, mellettük pedig az intézmény, település és ország neve került feltüntetésre. A folyamat eredményeként minden szerzőre ismertté vált az intézménynév, településnév és ország neve. Az adatok felhasználásakor a már validált és tisztított AU mező szerzői neveinek összekapcsolása történt a C1 mezőből kinyert szerzői nevekkal.

A WoS-ből származó rekordok feldolgozását követően három tábla állt elő, mely a különböző távolságok meghatározásának a kiindulópontját jelentette. Az első tábla a validált szerzői neveket és a cikkazonosítókat tartalmazta. Ebből hoztam létre a társszerzői mátrixot (6.1 fejezet), amely a társadalmi távolság azonosításához volt szükséges. A validált szerzői neveket és hivatkozásokat tartalmazó tábla a teljes kognitív távolság meghatározásának volt a feltétele. A validált szerzői neveket és szerzői affiliációkat tartalmazó tábla pedig a földrajzi távolság előző két távolsághoz való viszonyulásának vizsgálatához volt elengedhetetlen. A következő fejezetben ismertetem a konkrét módszereket és az általam elkülönített három fő távolságtípus egymáshoz való viszonyát.

6. Távolságtípusok meghatározása

A 4. fejezetben bemutattam, hogy hogyan értelmezhető a távolság fogalma társszerzői hálózatokon, illetve a vonatkozó szakirodalom alapján milyen távolságtípusok különíthetők el. Láttuk, hogy a vizsgált távolságtípusok hányféleképpen értelmezhetők, foghatók meg, és ennek megfelelően milyen sokféle módszer alkalmazható azonosításukra, mérésükre. Ugyanakkor több vizsgálat (pl. White et al., 2004; Wuyts et al., 2005; Ponds et al., 2007) arra is rámutatott, sokszor milyen nehéz elkülöníteni egymástól ezeket a távolságtípusokat, ugyanis ezek valójában folyamatosan összefonódnak, hatnak egymásra.

Ebben a fejezetben az előzőekben bemutatott adatbázison értelmezem az általam kiválasztott távolságtípusokat. Bemutatom, milyen módon, milyen módszerekkel azonosítottam a társadalmi, a teljes és tiszta kognitív, valamint a földrajzi távolságot. Fontos a módszerek körülhatárolása, hiszen az eddigiekben láthattuk, ugyanazon jelenség hányféleképpen ragadható meg a választott mérési módtól függően.

6.1. Társadalmi távolság meghatározása

A társadalmi távolság értelmezése, mint az a 4.2. fejezetben bemutatásra került, sokféleképpen történhet. Doktori értekezésem empirikus részében a társadalmi távolságot a társszerzői kapcsolatokon keresztül azonosítottam. A szerzők közötti társszerzői kapcsolatok alapján egy hasonlósági mátrixot határoztam meg. A hasonlósági mátrixon keresztül természetesen a két szereplő közötti távolság ugyanúgy mérhető, mintha távolságmátrixot hoztam volna létre, hiszen a két érték egymásba átfordítható (5.3. fejezet). A hasonlósági mátrix esetében minél magasabb érték szerepel két szerző között, annál kisebb közöttük a távolság.

Az 5.3. fejezetben bemutattam, hogy milyen konkrét adatokat használtam fel a vizsgálatomhoz. A társszerzői hasonlósági mátrix meghatározásához a WoS adatrekordokból az azonosított tisztított nevekre (AU) és a cikkek azonosítójára (UT) volt szükségem. Ebből a 4. táblázat felépítését követő táblát hoztam létre, ez lett a társszerzői mátrix alapja. Jelen példában van 5 szerző ($A1-A5$) és 5 publikáció ($P1-P5$). $P1$ cikket $A1$ és $A5$ szerző írta közösen, míg $P3$ cikk kizárólag az $A2$ szerzőhöz köthető. $P1$, $P2$ és $P5$ cikkeknek két-két társszerzője volt, míg $P4$ cikknek három. Ebből a táblából a 5. táblázat mutatja a nyers szerző-szerző társszerzőségi mátrixot. A szerző-publikáció mátrixból a szerző-szerző társszerzői mátrixot a következő formulával lehet megkapni:

$$M_{AA} = M_{AP} \times M_{AP}^T$$

ahol M_{AA} a szerző-szerző mátrix, M_{AP} pedig a szerző-publikáció mátrix. Az 5. táblázat sorai és oszlopai a szerzőket jelölik, míg a cellákban a közös publikációk száma szerepel. Az átlóban az egy szerző által írt összes publikáció száma található. A 4. táblázat alapján P1 cikk társszerzői az 5. táblázatban úgy jelennek meg, hogy A1 és A5 szerző között a cellaérték 1. (Ebből az is következik, hogy A1-nek és A5-nek más közös publikációja nincsen.)

	P1	P2	P3	P4	P5
A1	1	1	0	0	0
A2	0	0	1	1	0
A3	0	1	0	1	1
A4	0	0	0	1	0
A5	1	0	0	0	1

4. táblázat: szerző-publikáció mátrix

	A1	A2	A3	A4	A5
A1	2	0	1	0	1
A2	0	2	1	1	0
A3	1	1	3	1	1
A4	0	1	1	1	0
A5	1	0	1	0	2

5. táblázat: nyers társszerzői mátrix

A 6. táblázat a társszerzői hálózatnak azt a változatát mutatja, amikor a társszerzői kapcsolatokat a Salton-féle koszinusz hasonlósággal határoztam meg. A Salton-féle koszinusz hasonlóságot a tudományometriában leggyakrabban az együtt hivatkozások vizsgálatakor használják (Salton és McGill, 1983; Hamers et al., 1989), képlete a következő (Luukkonen et al., 1993):

$$S_{xy} = \frac{C_{xy}}{\sqrt{C_x \times C_y}}$$

ahol C_{xy} jelen esetben x és y szerző közös publikációinak a száma, C_x x szerző, C_y pedig y szerző összes publikációjának száma. S_{xy} értéke 0 és 1 között lehet. A Salton-féle koszinusz hasonlóság nem értelmezhető, ha $C_x = 0$ vagy $C_y = 0$, tehát x vagy y szerzőnek nincsen publikációja. Teljes hasonlóságról beszélünk, $S_{xy} = 1$, ha $C_{xy} = C_x = C_y$, azaz x és y szerző összes publikációja közös. Hogyha $C_{xy} = 0$, tehát x és y szerzőnek nincsen közös publikációja, nincsen hasonlóság, S_{xy} 0 értéket vesz fel (Vida, 2016).

A társadalmi hasonlósági mátrixot (M_s) bemutató 6. táblázat szerint példánkban A2 és A4 között találjuk a legerősebb hasonlóságot ($S_{A2A4} = 0,71$). Ennek háttérében az húzódik meg, hogy A4 szerző összesen egy publikációban vett részt, ekkor A2 és A3 szerzőkkel működött együtt. Ugyanakkor A2 szerző-

nek emellett csupán egy önálló publikációja volt, A3 szerző viszont mindenki mással is együttműködött, összesen három publikációban. A Salton-féle koszinusz hasonlóság indexe a társszerzős cikkek mellett a két szerző egyéb publikációinak a számát is figyelembe veszi.

	A1	A2	A3	A4	A5
A1	NA	0	0,41	0	0,50
A2	0	NA	0,41	0,71	0
A3	0,41	0,41	NA	0,58	0,41
A4	0	0,71	0,58	NA	0
A5	0,50	0	0,41	0	NA

6. táblázat: Salton-hasonlóságot tartalmazó társszerzői tábla

A bemutatott példához hasonlóan a közgazdaságtan és a természetföldrajz területén meghatároztam a vizsgált időszak publikációira a társszerzői Salton-féle koszinusz hasonlóság értékeit tartalmazó mátrixokat. A létrehozott hasonlósági mátrixokban 0 és 1 között szerepeltek értékek: a 0 azt jelentette, hogy a két szerző nem állt egymással társszerzői kapcsolatban a vizsgált időszakban, az 1 pedig azt, hogy a két szerző együttműködött valamennyi publikációjában. A magasabb értékek tehát erősebb társadalmi kapcsolatot jelöltek.

6.2. Kognitív távolság meghatározása

Az elemzés során a kognitív távolságot a közös hivatkozáshasználat alapján határoztam meg. Ehhez a bibliográfiai csatolás módszerét alkalmaztam, melynek értéke két dokumentum, szerző stb. közötti közös hivatkozások számával egyenlő (Ikpaahindi, 1986). A 11. ábra (4.3. fejezet, 73. old.) bal felső része a publikációk közötti bibliográfiai csatolás sematikus modellje. *P1* és *P2* publikáció az *R3* hivatkozáson keresztül áll kapcsolatban, tehát közöttük a bibliográfiai csatolás értéke 1.

Amennyiben két publikáció ugyanazon munkákra hivatkozik, feltételezhető, hogy hasonló kérdésekre keresik a választ, ily módon mögöttük hasonló tudásbázis áll, azaz kognitív kapcsolat állhat fenn közöttük. Ugyanakkor, amint arra Martyn (1964) és Tagliacozzo (1967) rámutatott, ez a kognitív kapcsolat csupán feltételezés marad, hiszen nem tudhatjuk, hogy a hivatkozott dokumentum ugyanazon megállapítására vonatkoznak-e a hivatkozó publikációk hivatkozásai. Azt, hogy a tématerületek lehatárolására a bibliográfiai csatolás mégis érvényes módszer lehet, Kessler (1965) a *Physical Review* folyóiratban megjelent publikációk csoportosításával bizonyította: a meglévő tárgyszóin-

dexelés, illetve a bibliográfiai csatolás módszere nagy átfedést mutató csoportokat alakított ki.

Az információtudomány, illetve a tudománymetria területén a bibliográfiai csatolás publikációk között történő alkalmazása bevett eljárás (Jarneving, 2007; Ahlgren és Jarneving 2008; Boyack és Klavans, 2010; Yan és Ding, 2012), azonban szerzői szinten ritkán használják. Példaként Zhao és Strotmann (2008a, 2008b) munkáit említhetjük, akik a publikációk szintjéről a szerzői szintre vetítették a hivatkozásokat (11. ábra jobb felső része). Ugyanakkor minden egyes publikáció esetén csak az első szerzőt vették figyelembe, azaz tulajdonképpen a publikációk adatait az első szerző nevére cserélték. Véleményem szerint a vetítés ily módon történő alkalmazásánál pont a lényeg veszik el a két szint között. Doktori kutatásomban a publikációk szintjéről a szerzői szintre történő vetítésnél ugyanazt a módot alkalmazom, mint amit Radicchi és munkatársai (2009) alkalmaztak az együtt hivatkozás szerzői szinten történő alkalmazásánál. Ők a folyóirat minden szerzőjéhez hozzárendelték az idézéseket, én is így jártam el, ahogy azt a 11. ábra bal és jobb alsó része mutatja a 4.3. fejezetben.

6.2.1. Teljes kognitív távolság meghatározása

A teljes kognitív távolság meghatározásához bibliográfiai csatolást alkalmaztam a szerzői szintre. Ehhez az 5.3. fejezetben bemutatott adatokból először szükségem volt a publikációk cikkazonosítójára (UT) és a 5.4.2. alfejezetben bemutatott, a vizsgálathoz előkészített hivatkozásokra (CR). Ezután a cikkekhez tartozó tisztított szerzői nevekhez (AU) rendeltem a hivatkozásokat, így egy szerző-hivatkozás mátrixot kaptam eredményül. Ebből, a társadalmi távolság méréséhez hasonlóan, itt is meghatároztam a szerző-szerző mátrixot a következő formula felhasználásával:

$$M_{AA} = M_{AR} \times M_{AR}^T$$

ahol M_{AR} a szerző-hivatkozás mátrix, M_{AA} pedig a szerző-szerző mátrix. Ez a szerzőpárok közös hivatkozásainak számát tartalmazta, azaz a szerzői bibliográfiai csatolást. Ahhoz, hogy a társszerzői hálózattal ez összehasonlítható legyen, a szerzők közötti bibliográfiai csatoláson alapuló hasonlóság meghatározásához is a Salton-indexet alkalmaztam. Így eredményül megkaptam a teljes kognitív hasonlósági mátrixot (M_{rc}), mely 0 és 1 közötti értékeket tartalmazott. A 0 a teljes különbséget, az 1 pedig a teljes egyezést jelentette két szerző között a közös hivatkozások alapján.

Látható, hogy az eljárás gyakorlatilag megegyezik a társadalmi távolság meghatározásánál alkalmazott módszerrel, hiszen mindkét esetben szerzőpárok kapcsolatának erősségét kerestem: az előző alfejezetben a társszerzői kapcsolatok, ebben az alfejezetben pedig a közös hivatkozások mentén. Így a 4., 5., 6. táblázatokban bemutatott példa erre a módszerre is érvényes lehet, azaz a különbséggel, hogy a 4. táblázat oszlopaiban publikációk helyett hivatkozások szerepelnek.

6.2.2. Tiszta kognitív távolság meghatározása

Az előző alfejezetben bemutatott teljes kognitív távolság két okból tartalmazhatott két szerző között kapcsolatot a közös hivatkozásokon keresztül. Az egyik esetben a társszerzőség miatt hivatkoztak ugyanazon publikációkra az egyes szerzők, a másikkban pedig a munkájukhoz használták ugyanazt a tudásbázist, így kognitív értelemben voltak közel egymáshoz. Mivel a teljes kognitív mátrix a két esetet együtt tartalmazta, az így meghatározott távolságnak a teljes kognitív távolság elnevezést adtam. Azonban a két esetet szerettem volna elkülöníteni egymástól. A 11. ábra (4.3. fejezet, 78. old.) jobb alsó része mutatta be azt a konceptuális modellt, mely a teljes kognitív távolság ketts tartalmát szemlélteti. A teljes kognitív távolságból a társszerzőség miatti kognitív kapcsolatok kiszűrését követően egy olyan mátrixot kapunk, amely a kognitív hasonlóság társszerzőségen kívül eső részét tartalmazza. Ezt tiszta kognitív hasonlósági mátrixnak (M_{PC}) neveztem el, és a teljes kognitív hasonlósági mátrix, valamint a társadalmi hasonlósági mátrix különbségének abszolút értékeként határoztam meg:

$$M_{PC} = |M_{FC} - M_S|$$

ahol M_{PC} a tiszta kognitív, M_{FC} a teljes kognitív, M_S pedig a társadalmi hasonlósági mátrix. Az abszolút értékre amiatt volt szükség, hogy a tiszta kognitív hasonlóságmátrix értékei is 0 és 1 közé essenek és az M_{PC} ne vegyen fel negatív értékeket. Vizsgálatomban az egyes hasonlósági értékek voltak számomra lényegesek, nem pedig a kapcsolatok iránya. A kapott tiszta kognitív hasonlóságmátrixban két szerző (A1 és A2) között 0 érték szerepelt $M_{PC1,2} = 0$, ha azok között nem volt közös hivatkozás $M_{FC1,2} = 0$, ekkor a két szerzőnek közös publikációja sem volt ($M_{S1,2} = 0$). Szintén $M_{PC1,2} = 0$, ha a két szerző minden cikke közös volt, azaz társszerzőségből adódott, ebben az esetben $M_{FC1,2} = M_{S1,2}$. A tiszta kognitív hasonlósági mátrixban két szerző között 1, azaz a maximum akkor lehetett a hasonlóság ($M_{PC1,2} = 1$), amikor A1 és A2 szerzőnek minden hivatkozása közös volt ($M_{FC1,2} = 1$), de nem publikáltak egy cikket sem közösen ($M_{S1,2} = 0$).

6.3. Földrajzi távolság vizsgálatának módja

A tudománymetriában a *térképezés* kifejezés alkalmazása kétféleképpen jelenik meg a szakirodalomban (Leydesdorff és Persson, 2010). Egyrészt *tudománytérképezés* révén a kognitív tématerületek egymáshoz való viszonyát ábrázolják ún. tudománytérképeken (Boyack et al., 2005; Leydesdorff és Rafols 2009; Rafols et al., 2010; Leydesdorff et al., 2013), másrészt a hagyományos *földrajzi térképezés* révén az együttműködési hálózatokat a földrajzi térben ábrázolják (Ponds et al., 2007; Leydesdorff és Persson, 2010). Ily módon lehetőség nyílik a térbeli dinamika elemzésére. Narin (1976) elsőként különböztette meg az országokat és a tudományterületeket, mint az értékelés fontos alapvető tényezőit, Small és Garfield (1985) pedig ezt a két közelítésmódot a térképezés különböző alapjaiként említette. Ez a két fajta közelítésmód a regionális tudomány belső tér – külső tér dichotómiájában úgy értelmezhető, hogy a tudománytérképezés a belső tér kapcsolatait írja le, míg a földrajzi térképezés a külső térben jeleníti meg a belső térben létrejövő kapcsolatokat.

Leydesdorff és Persson (2010) tanulmányukban a társszerzői együttműködési hálózatokat ábrázolták települési szinten, mellyel céljuk az SCI adatai és a térképi ábrázolás összekapcsolása volt. Ezenkívül online szabadon elérhető szoftverek alkalmazásával mutatták be a társszerzői hálózatok Google Térkép, illetve Google Earth segítségével történő megjelenítését. Szintén egy hálózat térképre vetítésére példa, amikor Lengyel és munkatársai (2013, 2015, 2016) egy online kapcsolatháló, az IWiW adatait elemezték. Munkájuk egyik fontos megállapítása, hogy az online közösségi hálózatok sem térfüggetlenek, a földrajzi tér hatással van alakulásukra. Jakobi (2016) többek között kiemelte, hogy a térinformatika nyújtotta vizualizációs eszközök hozzájárulhatnak a belső tér, így például az online baráti kapcsolathálózat külső térbeli leképeződésének vizsgálatához, jobb megértéséhez.

Doktori értekezésemben a fejezet elején bemutatott két típus közül a földrajzi térképezést alkalmaztam, mely során a teljes és tiszta kognitív, valamint a társadalmi hálózatokat a valós földrajzi térben vizsgáltam. A földrajzi távolságot a szerzők affiliációin keresztül elemeztem. A társszerzős publikációk affiliációinak vizsgálatakor problémát jelent, hogy az egyes szerzőkhöz tudjuk kötni a publikáción feltüntetett intézményi információkat. Gyakran előfordul, hogy egy társszerzős cikk szerzői között több szerző is egy intézményt tüntet fel, illetve egy szerző több intézményt is megad a publikáción. Katz és Martin (1997) számos példán keresztül szemléltette, hogy egy adott cikknek csupán az affiliációs adatai alapján nem kaphatunk pontos képet arról, hogy a cikket hány szerző jegyzi.

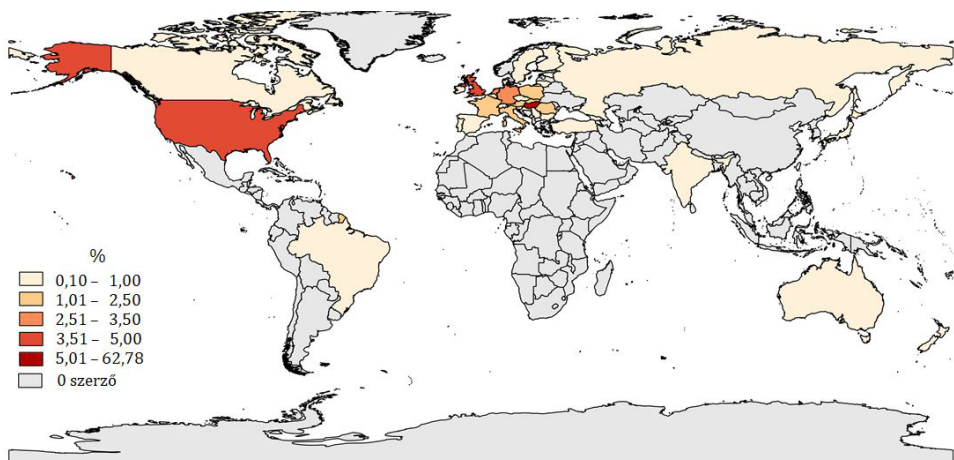
Az elemzésekkor egy-egy szerző esetében több affiliáció figyelembevétele egyre inkább elterjedt módszer a tudományometriában (Persson et al., 2004; Leydesdorff és Persson 2010). Ugyanakkor ez akár torzításokhoz is vezethet, például ha egy szerző két intézményt ad meg affiliációként és a vizsgálat során mindkettőt figyelembe veszik, akkor a települési szintű társszerzői hálózat megjelenítésekor e szerző két intézménye között szerepelni fog él. Ez félrevezető lehet, hiszen ilyenkor egy látszólagos együttműködési kapcsolat jön létre. Ugyanakkor onnan is meg lehet közelíteni ezt a kérdést, hogy a szerző többszörös affiliációja révén kapcsolatot teremt a két település között, még ha ez nem is nevezhető együttműködésnek. Tehát elmondható, hogy nézőponttól, kutatási céltól függően mindkét felfogásnak van létjogosultsága. Vizsgálatom során a többszörös affiliációval rendelkező szerzőknél az először előforduló intézmény települését vettem figyelembe, így a megjelenített térképeken nem valós együttműködési kapcsolat nem került ábrázolásra.

A hálózatok települési szintre történő vetítéséhez és térképen történő ábrázolásához először szükségem volt a szerzői affiliációk feldolgozására. A WoS rekord C1 mezője az adott publikáció minden egyes szerzőjéhez köthető módon tartalmazza az intézményi affiliációkat. Ez a földrajzi bibliometriai adatok elemzésénél nagyon lényeges. Az 5.4.2. alfejezetben részletesen leírtam, hogy a letöltött WoS rekordok C1 mezőjét hogyan készítettem elő az elemzéshez. A C1 mező feldolgozását követően, a vizsgált időszakban minden, a hasonlósági mátrixokban szereplő szerzőhöz hozzá tudtam rendelni az általa a publikációkon megadott affiliációs adatokat. Ez szerzőnként tartalmazta a szerző intézményének, az intézmény településének és országának a nevét. A különböző szintű földrajzi adatokból a települési szintre vonatkozó adatokat használtam fel a doktori kutatásban. A későbbi egyszerűbb szűrési lehetőségekhez különböző dummy változókat vezettem be: egy adott szerző intézménye Magyarországon található-e (HU), Európában található-e (EU), Magyarországgal szomszédos országban található-e (HUszomszéd). A 7. táblázatban néhány példán keresztül láthatjuk az így előállt tábla felépítését.

Szerző azonosító	Affiliáció	Településnév	Ország	HU	EU	HUzsomszéd
21	Univ Szeged, Fac Econ & Business Adm, Szeged, Hungary	Szeged	Hungary	1	1	1
22	Szechenyi Istvan Univ, Dept Logist & Forwarding, Gyor, Hungary	Gyor	Hungary	1	1	1
23	Szent Istvan Univ, Inst Engn Management, H-2103 Godollo, Hungary	Godollo	Hungary	1	1	1
24	Santa Clara Univ, Food & Agribusiness Inst, Santa Clara, CA 95053 USA	Santa Clara	USA	0	0	0
25	Corvinus Univ Budapest, Dept Microecon, H-1093 Budapest, Hungary	Budapest	Hungary	1	1	1
26	Hungarian Acad Sci, Inst Econ, H-1112 Budapest, Hungary	Budapest	Hungary	1	1	1
28	Szent Istvan Univ, Environm Social Sci Res Grp, Szent Istvan, Hungary	Godollo	Hungary	1	1	1
29	Univ W Hungary, Fac Econ, Sopron, Hungary	Sopron	Hungary	1	1	1
30	Univ Szeged, Dept Med 1, H-6720 Szeged, Hungary	Szeged	Hungary	1	1	1
35	Cent European Univ, Dept Econ, H-1051 Budapest, Hungary	Budapest	Hungary	1	1	1
36	Corvinus Univ Budapest, H-1093 Budapest, Hungary	Budapest	Hungary	1	1	1
37	Szent Istvan Univ, Fac Econ & Social Sci, Budapest, Hungary	Budapest	Hungary	1	1	1
40	Univ Durham, Durham Business Sch, Durham DH1 3HY, England	Durham	England	0	1	0
41	Cent European Univ, Dept Publ Policy, H-1051 Budapest, Hungary	Budapest	Hungary	1	1	1
42	Univ Babes Bolyai, Fac Business, R-3400 Cluj Napoca, Romania	Cluj Napoca	Romania	0	1	1
43	IE HAS, H-1112 Budapest, Hungary	Budapest	Hungary	1	1	1
44	Univ Pecs, Fac Business & Econ, Pecs, Hungary	Pecs	Hungary	1	1	1
45	Natl Res Univ, Higher Sch Econ, ICEF, Moscow, Russia	Moscow	Russia	0	1	0
46	Dept Econ & Law, I-00161 Rome, Italy	Rome	Italy	0	1	0

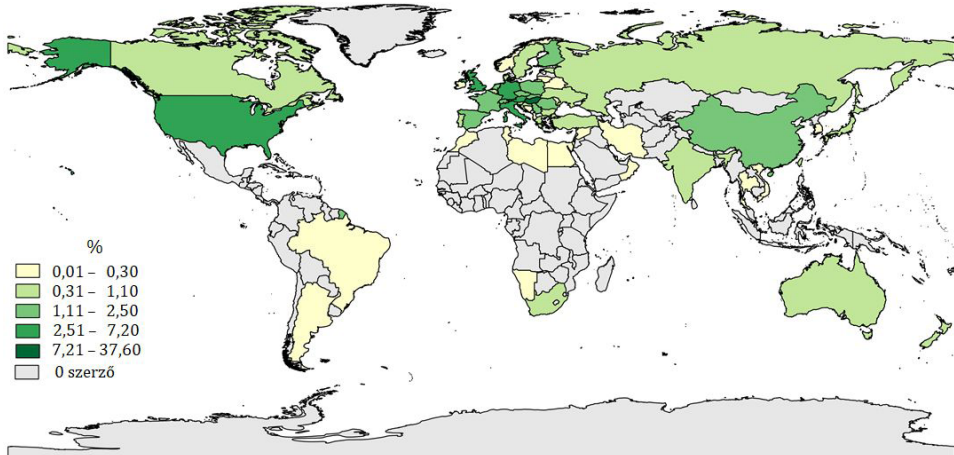
7. táblázat: Szerzők affiliációinak azonosítását követően előállt tábla példa

A 17. és 18. ábra térképen mutatja a vizsgálatba vont publikációk szerzőinek országok szerinti megoszlását. Mind a két tudományterületen a legtöbb szerző magyarországi affiliációval rendelkezett (közgazdaságtan: 62,78%, természetföldrajz: 37,6%). Ez az adatgyűjtésből adódott, hiszen a legalább egy magyar affiliációval rendelkező publikációk kerültek be a vizsgálatba (5.3. fejezet). A térképeken megfigyelhetjük Magyarország fő társszerzői partnerországait 2010 és 2014 között a közgaz-



17. ábra: A mintában szereplő szerzők megoszlása országok szerint a közgazdaságtan területén

daságtan és a természetföldrajz területén. A mintában megjelenő fő társ-szerzői partnerországok megegyeznek Magyarország teljes tudományos kibocsátására jellemző fő partnerországaival (MTA KIK TTO jelentések, 2013, 2014, 2015, 2016, 2017). Ez alátámasztja a minta megbízhatóságát.



18. ábra: A mintában szereplő szerzők megoszlása országok szerint a természetföldrajz területén

A földrajzi távolság vizsgálatokor a korábban (6.1 és 6.2. fejezet) meghatározott társadalmi, teljes és tiszta kognitív hasonlósági mátrixok hálózatként való értelmezését használtam ki. Mindhárom hasonlósági mátrix megad egy hálózatot, melyben a csúcsok a szerzők, az élek pedig a hasonlósági mátrix cellaértékei. Ahhoz, hogy ezeket kifeszítsem a földrajzi térbe, a szerzők települési affiliációit vettem figyelembe, azaz a hálózati csúcsok a szerzők helyett a szerzők települési affiliációival lettek helyettesítve: így váltak láthatóvá a szerzők közötti kapcsolatok a települések között. Ehhez először a szerzői hálózatok kapcsolatait éllista formájában határoztam meg (8. táblázat), majd a szerzőkön keresztül az affiliációkat tartalmazó táblával kapcsoltam össze (9. táblázat). Így egy olyan hálózatot kaptam, ahol a csúcsok a települések, az élek pedig a társadalmi, valamint a teljes és tiszta kognitív hálózatok települések között megjelenített kapcsolatait voltak.

from	to	weight
1	3	0,0111
1	10	0,0174
1	15	0,0251
1	43	0,0147
1	59	0,0058
1	115	0,0121
1	250	0,0218
1	300	0,0174
1	445	0,0174
1	567	0,6774

8. táblázat: Példa szerzői hálózatok éllistájára

Megjegyzés: a from és to oszlopok a csúcsokat jelölik és a szerzők azonosítóit tartalmazzák; mivel nem irányítottak a hálózatok, ezek az oszlopok fel is cserélhetők. A weight oszlop az élsúly értékét tartalmazza.

from	from.city	from.lat	from.long	from.co	to	to.city	to.lat	to.long	to.co	weight
2	Palmerston North	-40.352306	175.608215	New Zealand	100	Frankfurt	50.110922	8.682127	Germany	0,009527699
4	Budapest	47.497912	19.040235	Hungary	100	Frankfurt	50.110922	8.682127	Germany	0,006517182
30	Budapest	47.497912	19.040235	Hungary	100	Frankfurt	50.110922	8.682127	Germany	0,012320355
70	Vienna	48.208174	16.373819	Austria	100	Frankfurt	50.110922	8.682127	Germany	0,013608276
89	Budapest	47.497912	19.040235	Hungary	100	Frankfurt	50.110922	8.682127	Germany	0,073989544
140	Warsaw	52.229676	21.012229	Poland	100	Frankfurt	50.110922	8.682127	Germany	0,007607258
210	Helsinki	60.169856	24.938379	Finland	100	Frankfurt	50.110922	8.682127	Germany	0,047740992
305	Helsinki	60.169856	24.938379	Finland	100	Frankfurt	50.110922	8.682127	Germany	0,047740992
400	Palmerston North	-40.352306	175.608215	New Zealand	100	Frankfurt	50.110922	8.682127	Germany	0,009527699
460	Okayama	34.655146	133.919502	Japan	100	Frankfurt	50.110922	8.682127	Germany	0,010758287

9. táblázat: Példa affiliációkat is tartalmazó éllistára

Megjegyzés: A from és to oszlopok az él két végpontját különítik el. Irányítatlan hálózat lévén ezek akár fel is cserélhetők. from vagy to: szerzői azonosító él egyik vége, from.city vagy to.city: településnév él egyik vége, from.lat vagy to.lat és from.long vagy to.long: településnév él egyik végének szélességi és hosszúsági koordinátái, from.co. vagy to.co: országnév él egyik vége, weight: élsúly.

wkt	weight
LINESTRING(175.608215 -40.352306,8.682127 50.110922)	0,009527699
LINESTRING(19.040235 47.497912,8.682127 50.110922)	0,006517182
LINESTRING(19.040235 47.497912,8.682127 50.110922)	0,012320355
LINESTRING(16.373819 48.208174,8.682127 50.110922)	0,013608276
LINESTRING(19.040235 47.497912,8.682127 50.110922)	0,073989544
LINESTRING(21.012229 52.229676,8.682127 50.110922)	0,007607258
LINESTRING(24.938379 60.169856,8.682127 50.110922)	0,047740992
LINESTRING(24.938379 60.169856,8.682127 50.110922)	0,047740992
LINESTRING(175.608215 -40.352306,8.682127 50.110922)	0,009527699
LINESTRING(133.919502 34.655146,8.682127 50.110922)	0,010758287

10. táblázat: Példa hálózat éleinek térképi vonalobjektumként történő megadására

Megjegyzés: a wkt a vonalobjektum, weight: élsúly.

A térképi megjelenítéshez a QGIS open source szoftvert használtam. A hálózatok alap shape fájlban történő megjelenítéséhez az éleket vonalobjektumként kellett meghatároznom. Ehhez először a www.doogal.co.uk/BatchGeocoding.php honlap segítségével meghatároztam a települések koordinátáit, melyek egyben a vonalak kezdő- és végpontjait is jelentették. Ezek az alap shape fájlal megegyező vetületben voltak, így a települések koordinátái is azonosak voltak.

Ezt követően azt kellett meghatároznom, mely települések között van kapcsolat; ezt az egyes hálózatok kapcsolatai határozták meg. Mindegyik hálózatot éllistaként adtam meg, ahol a szerzők közötti élek kerültek megadásra, ezt követően ehhez rendeltem hozzá a szerzői affiliációkat tartalmazó táblát, mellyel egy települések közötti éllistát hoztam létre. A települési pontkoordinátákból a kapott éllista alapján vonalobjektumokat hoztam létre, a következő formában megadva az éllista koordinátáit: LINESTRING(*fromx_fromy*, *tox_toy*); ahol *fromx* és *fromy* az él egyik végpontjának koordinátáit, míg *tox* és *toy* a másik végpont koordinátáit jelöli (10. táblázat). A *from* és *to* elnevezés megtévesztő lehet, de mivel irányítatlan hálózatokat ábrázoltam, természetesen az élek végpontjainak koordinátái akár fel is cserélhetők egymással. Az így létrehozott térképen megjelenített két tudományterület három-három hálózatát a 7.3. fejezetben hasonlítom össze.

7. A létrehozott távolságtípusok összehasonlítása

Az előző, 6. fejezetben a három vizsgált távolságtípus hasonlósági mátrixokkal történő meghatározásának pontos menete került leírásra. Mindkét tudományterületre (közgazdaságtan és természetföldrajz) létrehoztam a szerzők közötti közelséget leíró mátrixokat, melyek teljesen egyforma méretűek, mivel ugyanazon szerzők szerepelnek bennük, ABC-sorrendben. Ezek alapján először a társadalmi és kognitív távolság összehasonlítását végeztem el. Ez tulajdonképpen három-három mátrix összehasonlítását jelenti: ezek a teljes kognitív távolság, valamint komponensei, a társadalmi és a tiszta kognitív távolság mátrixai.

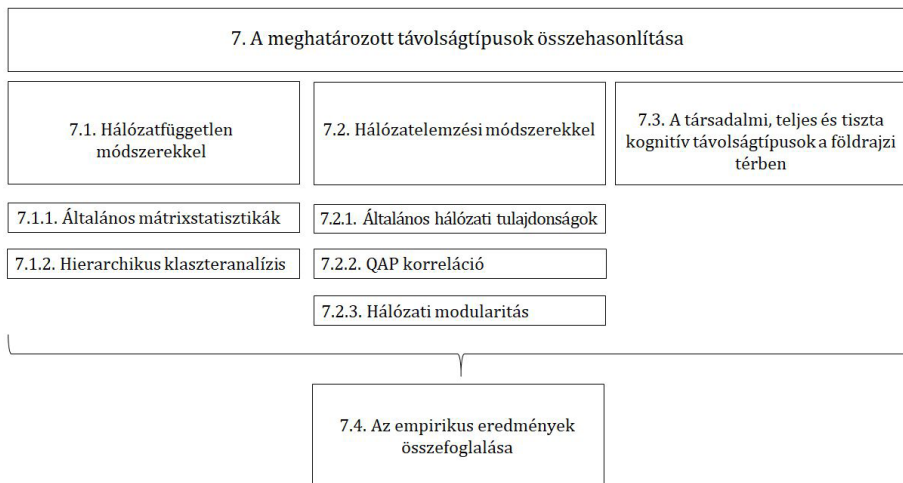
A meghatározott távolságtípusokat a hasonlósági mátrixokon keresztül vettem össze. Elemzési módszereim két fő csoportra bonthatók: az egyikben (7.1. fejezet) a létrehozott mátrixok jellemzőit hálózathoz nem kötött módon, *általános mátrixstatisztikák* (7.1.1. alfejezet), valamint *klaszteranalízis* (7.1.2. alfejezet) által létrejött tématerületek összehasonlítása segítségével végeztem. A klaszteranalízis itt azt a célt szolgálta, hogy megmutassa, a szerzői csoportok által meghatározott tématerületek mutatnak-e érdemi különbséget a különböző távolságtípusok között.

A másik fő irányvonalban (7.2. fejezet) a hasonlósági mátrixok hálózatként történő felfogása került előtérbe, így azokat különböző hálózatelemzési módszerekkel is összehasonlíthattam. Mindegyik hasonlósági mátrix megad egy hálózatot, ahol a csúcsok a szerzők, az élek pedig a szerzők közötti közelségértékek. Ezeket a hálózatokat először *általános hálózati tulajdonságok* (7.2.1. fejezet) mentén jellemeztem: olyan mutatókon túl, mint például a csúcsok és élek száma, a sűrűség, a globális klaszterezettség aránya, a hálózati távolságok is összehasonlításra kerültek. Módszertani okokból ez utóbbiakat csak a csúcsok többségét magába foglaló óriáskomponensen belül elemeztem. A következő módszer az ún. *Quadratic Assignment Procedure, QAP korreláció* (7.2.2. fejezet). Ez a hálózatok szerkezeti hasonlóságát méri, megmutatja annak valószínűségét, hogy ha az egyik hálózatban két szereplő között van kapcsolat, ugyanazon a helyen a másik hálózatban is találunk élt. Végül a *hálózati modularitást* (7.2.3. fejezet) is kiszámítottam, fast greedy algoritmus segítségével. Ez az algoritmus modulokat határoz meg a hálózatokon belül: az egy hálózati modulba tartozó csúcsok erősebben kapcsolódnak egymáshoz, mint a modulhoz nem tartozó csúcsokhoz (Csermely, 2005).

A hasonlósági mátrixok összehasonlítását végül a földrajzi távolság „bekapcsolásával” együtt is elvégeztem. Ekkor a 6.3. fejezetben leírtak szerint a szerzői hasonlósági mátrixok által meghatározott hálózatokat a települési szintre

aggregáltam, és így lehetőségem adódott a társadalmi, teljes kognitív és tiszta kognitív hálózatok földrajzi térben történő összehasonlítására (7.3. fejezet).

A meghatározott hasonlósági mátrixok összehasonlításának módjait a következő folyamatábrán foglalom össze (19. ábra).



19. ábra: A 7. fejezet felépítése

7.1. Távolságtípusok összehasonlítása hálózatfüggetlen módszerekkel

7.1.1. A társadalmi és a kognitív hasonlóságmátrixok kapcsolata

A mátrixokat először a hasonlóság erősségének szintjein vetettem össze. A 11. táblázat azt mutatja, hogy a kialakított mátrixokban az összes mező hány százaléka esett az egyes erősségi tartományokba.

	Közgazdaságtan			Természetföldrajz		
	Teljes kognitív hasonlóság (%)	Társadalmi hasonlóság (%)	Tiszta kognitív hasonlóság (%)	Teljes kognitív hasonlóság (%)	Társadalmi hasonlóság (%)	Tiszta kognitív hasonlóság (%)
>=0,1	0,62637	0,54353	0,18347	0,71854	0,62101	0,22707
>=0,2	0,54272	0,53504	0,03556	0,62572	0,61618	0,06756
>=0,3	0,50999	0,50716	0,00889	0,58580	0,60105	0,00445
>=0,4	0,46877	0,47362	0,00566	0,56690	0,58375	0,00091
>=0,5	0,42917	0,43806	0,00485	0,54527	0,56512	0,00008
>=0,6	0,38876	0,37259	0,00040	0,52113	0,50466	0,00000
>=0,7	0,35198	0,36936	0,00000	0,48767	0,50109	0,00000
>=0,8	0,31804	0,29258	0,00000	0,44413	0,40638	0,00000
>=0,9	0,29177	0,28530	0,00000	0,39501	0,37489	0,00000

11. táblázat: Hasonlósági mátrixok összehasonlítása kapcsolaterősségi szintenként

A hasonlósági mátrixokat ezen erősségi szintek mentén vizsgáltam. Mindezenelőtt meg kell jegyeznünk, hogy valamennyi mátrix mezőinek döntő többsége 0 értéket vett fel, azaz az adott szerzőpárok között a kérdéses távolságtípus dimenziójában nem volt kapcsolat. Ez magyarázza azt is, hogy a 11. táblázatban miért szerepeltek ilyen alacsony értékek.

Ha a két vizsgált tudományterületet a teljes kognitív és a társadalmi hasonlóság kapcsolaterősségi szintjei szerint vetjük össze, láthatjuk, hogy a természetföldrajz területén rendre több kapcsolat esik az egyes kapcsolaterősségi kategóriákba, azaz ezen a tudományterületen erősebb kapcsolatok találhatók a szerzők között, mint a közgazdaságtan esetében. Ez magyarázható azzal a már említett ténnyel, hogy a természettudományok területén gyakoribbak a társszerzőségek. A teljes kognitív mátrix (a 4.3. fejezetben bemutatott konceptuális modell alapján) a társszerzőségek révén tartalmazza a társadalmi hasonlóságot is. Az eredményekből az látszik, hogy a teljes kognitív mátrixon belül a társadalmi komponens nagyon domináns, így a természetföldrajz területén jelenlevő magasabb társszerzői arány itt is erősebb kapcsolatokat eredményez. Amikor azonban a társszerzőségi hatást kiszűrtem, a tiszta kognitív hasonlóság kapcsolaterősségi szintjein mást tapasztaltam: az alacsonyabb kapcsolaterősségi szinteken még a természetföldrajz területén találtam több kapcsolatot, 0,3-as szint fölött azonban a folyamat átfordult, a közgazdaságtan szerzői között nagyobb arányban voltak jelen az erősebb kapcsolatok. Ez azt jelenti, hogy a tiszta kognitív hasonlóság a közgazdaságtan területén volt erősebb a szerzők között. Meg kell azonban jegyezni, hogy összességében a tiszta kognitív hasonlósági mátrixban a kapcsolatok erőssége jóval alacsonyabb értékeket mutatott.

7.1.2. Hierarchikus klaszteranalízis által létrejött tématerületek összehasonlítása

A következőkben a különböző távolságtípusokat a klaszteranalízis által meghatározott témacsoportok elemzésén keresztül vizsgálom. A klaszteranalízis a tudománymetriában gyakran alkalmazott módszer arra, hogy a bibliográfiai csatolás, az együtt hivatkozás, illetve a kulcsszóelemzés eredményeinek csoportosításával tématerületek kerüljenek elkülönítésre (Ahlgren és Jarneving 2008; Janssens et al., 2009; Boyack és Klavans, 2010; Yan és Ding, 2012). A hierarchikus klaszteranalízis segítségével szerzői csoportokat hoztam létre. Ezt követően az egyes csoportokban található közös publikációkat vagy közös hivatkozásokat elemezve, a legjellemzőbb témák alapján neveztem el azokat. A kutatásomban nem a tématerületek megállapítása volt az elsődleges célom, e módszeren keresztül is a különböző távolságtípusokat hasonlítottam össze. A klaszteranalízis segítségével teszteltem, hogy érdemben különböznek-e egymástól a létrehozott távolságtípusok a szerzői csoportok által meghatá-

rozott tématerületek szerint, illetve a létrejött szerzői csoportok résztvevői mennyire fedik át egymást.

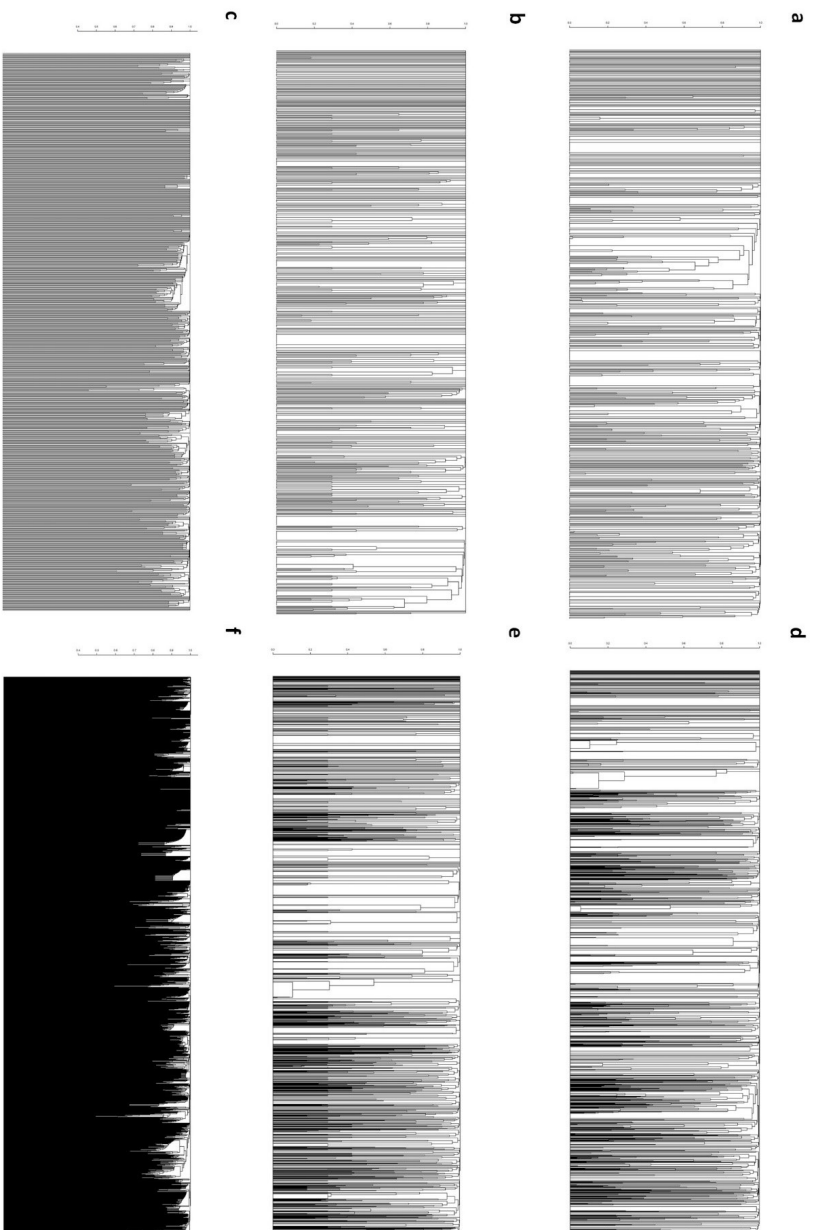
A hierarchikus klaszteranalízis bemeneti adatként távolságmátrixokat igényel, így a hasonlósági mátrixokból távolságmátrixokat hoztam létre a már ismertetett módon (5.3. fejezet). A távolságmátrixok szerzőit hierarchikus klaszterelemzéssel csoportosítottam mindkét tudományterületen. A teljes és tiszta kognitív távolságmátrix esetében a csoportok lehatárolása a közös hivatkozáshasználat alapján történt, a társadalmi távolságmátrix esetében pedig a közös cikkek jelentették a csoportosítás alapját.

Annak megállapításához, hogy a különböző hierarchikus klaszteranalízis-eljárások közül melyik kerüljön alkalmazásra, megvizsgáltam a kofenetikus korrelációs együtthatókat. A kofenetikus korreláció a kofenetikus távolságok és az eredeti távolságmátrix közötti kapcsolat erősségét állapítja meg. Két elem kofenetikus távolsága a dendrogram azon magassága, melyen a két elem egy klaszterbe kerül. A módszer annak megállapítására szolgál, hogy egy adott távolságmátrixra melyik klaszterelemzési algoritmus a legalkalmasabb (Tan et al., 2007). Az eljárással a Ward's, Single, Complete, Average, Mcquitty, Median, Centroid algoritmusokat hasonlítottam össze: mindegyik esetben az Average algoritmus kofenetikus korrelációs együtthatója volt a legmagasabb (12. táblázat), így ez bizonyult a legmegbízhatóbbnak a jelen adatokon.

	Közgazdaságtan			Természetföldrajz		
	Teljes kognitív távolságmátrix	Társadalmi távolságmátrix	Tiszta kognitív távolságmátrix	Teljes kognitív távolságmátrix	Társadalmi távolságmátrix	Tiszta kognitív távolságmátrix
Average	0,97	0,97	0,78	0,98	0,98	0,72
Mcquitty	0,96	0,96	0,73	0,97	0,97	0,63
Complete	0,95	0,95	0,48	0,97	0,96	0,33
Centroid	0,86	0,81	0,04	0,82	0,74	0,13
Single	0,72	0,69	0,47	0,43	0,34	0,29
Median	0,60	0,53	0,10	0,45	0,42	0,07
Ward.D	0,08	0,09	0,02	0,08	0,07	0,03

12. táblázat: A különböző klaszterezési algoritmusok kofenetikus korrelációs együtthatói

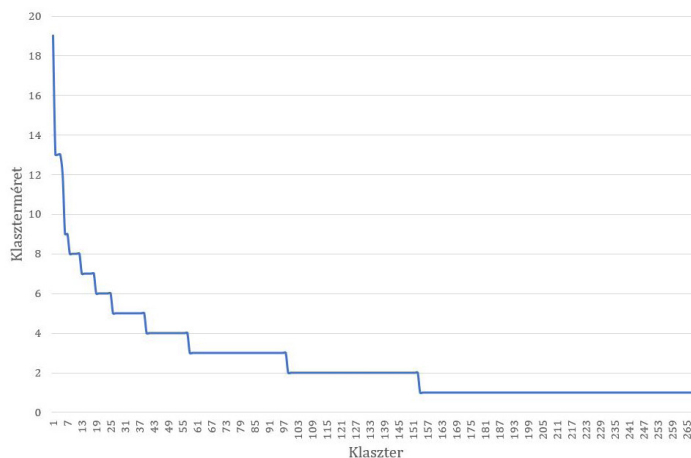
A 20. ábra a távolságmátrixok dendrogramjait mutatja. Megállapítható, hogy valamennyi esetben elég hamar kisebb csoportokra bomlott szét a dendrogram, már a 0,9-es vágási szinten is több száz klaszter alakult ki. Az is megfigyelhető, hogy a természetföldrajz területén mindig több klaszter képződik, mint a közgazdaságtan esetében.



20. ábra: Teljes kognitív (a,d), társadalmi (b,e) és tiszta kognitív (c,f) távolságmátrixok dendrogramjai a közgazdaságtan (a,b,c) és a természetföldrajz (d,e,f) területén

A klaszteranalízishoz használt R statisztikai szoftverben egy beépített ún. „dinamikus vágás” (*dynamicTreeCut*) függvény is található. Ez a dendrogram struktúráján felülről lefelé halad végig és a lehető legkevesebb, a szerkezet alapján egymástól elkülönülő klasztert jelöli ki. Ugyanakkor a függvény létrehoz egy ún. 0. klasztert is, mely gyakorlatilag a nagyon kis elemszámú vagy egyelemű klasztereket olvasztja össze egyetlen klaszterbe. A 0. klasztert emiatt nem is szükséges bevonni az értékelésbe. A *dynamicTreeCut* algoritmus jóval kevesebb számú klasztert hozott létre, mint a dendrogram egy adott szinten történő vágása. A természetföldrajz esetén a dinamikus vágás módszerével 30–40, míg a közgazdaságtan esetén csak néhány klaszter alakult ki: az is előfordult, hogy csak két klaszter keletkezett, melyből az egyik a már leírt 0. klaszter volt. Ez ilyenkor sajnos arra utalt, hogy nem különültek el a dendrogramban értékelhető módon a klaszterek. Bár a dinamikus vágás függvény a kis elemszámú klaszterek kiszűrése miatt jobb választásnak tűnhet, a vizsgált mintán nem mindig hozott megfelelő eredményt, ezért úgy döntöttem, hogy az értékelés során a fix vágási szinttel dolgozom.

A 21. ábra a 0,9-es vágási szinten kialakult klaszterek méretének eloszlását mutatja a közgazdaságtan területén a társadalmi távolságmátrix hierarchikus klaszterezését követően. A legnagyobb öt klaszter után egy nagyobb törés látható, így a klaszterek jellemzését csak a törés előtti csoportokra korlátoztam. A többi távolságmátrix 0,9-es szintű vágását követő klaszterméreteinek eloszlása a 21. ábrához hasonló képet mutat. A 13. táblázat a 0,9-es vágási szinten kialakult öt, illetve tíz legnagyobb klaszter elemszámának az összes szerzőhöz viszonyított arányát mutatja mátrixonként. A 0,9-es vágás alapján létrejött klaszterek közül tehát az öt legnagyobbat elemeztem, ezek együttesen, mátrixtól függően a szerzők 10–30%-át foglalták magukba.



21. ábra: Klasztercsoportok méretének eloszlása a társadalmi távolságmátrixon a közgazdaságtan területén

	Közgazdaságtan		Természetföldrajz	
	10 legnagyobb klaszter szerzőinek aránya az összes szerzőből (%)	5 legnagyobb klaszter szerzőinek aránya az összes szerzőből (%)	10 legnagyobb klaszter szerzőinek aránya az összes szerzőből (%)	5 legnagyobb klaszter szerzőinek aránya az összes szerzőből (%)
Teljes kognitív távolságmátrix	19	12	19	12
Társadalmi távolságmátrix	16	10	17	10
Tiszta kognitív távolságmátrix	33	30	34	21

13. táblázat: A hierarchikus klaszteranalízis legnagyobb tíz és öt klaszterének relatív mérete

Az így kialakított klasztercsoportokat először a módosított Rand-index segítségével vizsgáltam. Összevettem, hogy a szerzők klasztercsoportokba történő besorolása a különböző távolságtípusok esetén mennyire volt átfedésben. A Rand-index segítségével megállapítható, hogy egy adatsor kétféle csoportosítása mennyire hasonlít egymáshoz. Az index páronként összehasonlítja az egyes eseteket, és egy 0 és 1 közötti értéket ad eredményül. 0 esetén a két csoportosítás teljesen eltér egymástól, míg 1 esetén a két csoportosítás megegyezik (Rand, 1971). A Rand-index képlete a következő:

$$R = \frac{a + b}{n(n - 1)/2}$$

ahol a azon esetek száma, amikor a két csoportosításban megegyeztek az elemek, b azon esetek száma, amikor a két csoportosításban nem egyeztek meg az elemek, n pedig az elemszám. Hubert és Arabie (1985) megalkotta a Rand-index módosított változatát, mely a véletlenből származó egyezések és különbözőségek figyelembevételével korrigálja az eredeti indexet. Vizsgálatomban a módosított Rand-indexet használtam a létrejött szerzői klaszterek egyezőségének megállapítására.

	Teljes kognitív távolságmátrix	Társadalmi távolságmátrix	Tiszta kognitív távolságmátrix
Teljes kognitív távolságmátrix	1	0,8491	0,0133
Társadalmi távolságmátrix	0,8491	1	0,0104
Tiszta kognitív távolságmátrix	0,0133	0,0104	1

14. táblázat: A szerzők klasztercsoportokba történő sorolásának hasonlósága módosított Rand-index alapján a közgazdaságtan területén

	Teljes kognitív távolságmátrix	Társadalmi távolságmátrix	Tiszta kognitív távolságmátrix
Teljes kognitív távolságmátrix	1	0,8841	0,0933
Társadalmi távolságmátrix	0,8841	1	0,0790
Tiszta kognitív távolságmátrix	0,0933	0,0790	1

15. táblázat: A szerzők klasztercsoportokba történő sorolásának hasonlósága módosított Rand-index alapján a természetföldrajz területén

A 14. táblázat és a 15. táblázat a létrehozott szerzői klaszterek hasonlóságot vizsgálta a két tudományterületen. A teljes kognitív és a társadalmi távolságmátrix alapján létrejött szerzői klaszterek nagyfokú hasonlóságot mutatnak, míg a társadalmi és a tiszta kognitív, valamint a teljes és tiszta kognitív távolságmátrixok alapján meghatározott szerzői klaszterek nem voltak azonosak. Megállapítható, hogy a doktori kutatásomban definiált távolságtípusok jól elkülönülnek egymástól. A létrehozott tiszta kognitív távolság mentes a társadalmi távolság hatásától. A teljes kognitív és a társadalmi távolság hasonlósága a társszerzőségek miatt megvalósuló közös hivatkozások magas aránya miatt jelent meg.

Ezt követően a klaszterek létrejöttét meghatározó tényezők alapján jellemeztem a klasztereket, majd a bennük található hivatkozások és publikációk alapján elneveztem azokat tématerületek szerint.

A klaszterek létrejöttét meghatározó tényező – például a teljes kognitív távolságmátrix esetében a közös hivatkozáshasználat – alapján minden egyes klaszter esetében megnéztem, hogy az abban szereplő szerzők között melyek voltak a leggyakoribb hivatkozások.

Ahogy az 5. fejezetben kifejtettem, a WoS rekordok hivatkozási mezői tartalmazták a szerzők vezetéknéveit, a megjelenés évét, a folyóirat nevét, évfolyamát, számát, oldalszámát, néhol a DOI azonosítót (16. táblázat). Ezek közül a témára vonatkozóan csak a folyóirat neve volt közvetlenül használható. A klaszter témabesorolásához így a hivatkozásoknál található folyóiratok neveit használtam kumulált formában. A hivatkozásokból kinyertem a folyóiratokat és megnéztem, hogy az egyes folyóiratok milyen gyakorisággal fordultak elő az egyes klaszterekben.

CR mező	Folyóirat a CR mezőből
Hettrich S, 2014, ASTROBIOLOGY, V14, P377, DOI 10.1089/ast.2013.1076	ASTROBIOLOGY
Kereszturi A, 2011, ACTA ASTRONAUT, V68, P1686, DOI 10.1016/j.actaastro.2010.11.008	ACTA ASTRONAUT
Yingst RA, 2013, ACTA ASTRONAUT, V90, P311, DOI 10.1016/j.actaastro.2011.10.001	ACTA ASTRONAUT
Groemer G, 2014, ASTROBIOLOGY, V14, P391, DOI 10.1089/ast.2013.1081	ASTROBIOLOGY
Abramov I, 2005, ACTA ASTRONAUT, V57, P901, DOI 10.1016/j.actaastro.2005.05.001	ACTA ASTRONAUT
Bapna D, 1998, IEEE INT CONF ROBOT, P597, DOI 10.1109/ROBOT.1998.677038	IEEE INT CONF ROBOT
Belka Z, 1998, J SEDIMENT RES, V68, P368	J SEDIMENT RES

16. táblázat: Web of Science rekord Cited References (CR) mezője és abból a kinyert folyóirat

A folyóiratok alapján történő témabesorolásnak előnyei és hátrányai is vannak. Hátránya, hogy a folyóiratok címei csak hozzávetőleges információt adnak a felhasznált cikkek témájáról, így csak sokkal általánosabb tématerületek állapíthatók meg. A módszer előnye, hogy minden hivatkozásra alkalmazható volt, így nem csak a WoS-ban található hivatkozások alapján került elnevezésre az adott klaszter; az is a módszer mellett szólt, hogy további adatlekérdezés nélkül viszonylag pontos képet kaphattam a tématerületekről. A hivatkozások alapján részletesebb témabesorolást csak oly módon lehetett volna kivitelezni, ha minden egyes hivatkozást visszakeresek a WoS adatbázisban, és az eredeti forráscikkhez tartozó teljes WoS rekordból felhasználom a cikkek témabesorolását, címét és kulcsszavait. Ez valószínűleg pontosabb képet adott volna, viszont a hivatkozások közül csak a WoS-ban szereplő cikkeket lehetett volna bevonni az elemzésbe. Ez a magyar nyelvű cikkekre történő hivatkozások jó részét kizárhatta volna.

A társszerzői távolságmátrix klaszterezése a közös cikkek alapján történt. Ez az eleve letöltött WoS rekordokat jelentette, ahol rendelkezésre állt a folyóirat neve (SO) mellett a cím (TI) és szerzői kulcsszavak (DI) is. Megállapítottam, hogy a klaszterek téma alapján történő elnevezéséhez a folyóiraton kívül a cím és kulcsszó mező használata nem adott hozzá számottevő többletinformációt. Ez is alátámasztja azt, hogy a hivatkozási távolságmátrix alapján történő klaszterek témaelnevezéséhez elégséges volt a folyóiratok neveinek használata. Az egységes elnevezési eljárás végett a társadalmi távolságmátrix esetében is csak a folyóiratok kumulált klaszterenkénti gyakoriságát használtam.

A klaszterelemzés célja annak megállapítása volt, hogy vajon a kognitív és társadalmi távolság alapján kirajzolódó tématerületek különböznek-e egymástól. Ehhez mindhárom távolságmátrixon meghatároztam a folyóiratok

neveiből kirajzolódó fő témacsoportokat. Ez megfelelt egy általános kép felvázolásához.

Teljes kognitív távolságmátrix	Klaszter neve	Klaszter méret (%)	Klaszter neve	Klaszter méret (%)	Klaszter neve	Klaszter méret (%)	Klaszter neve	Klaszter méret (%)	Klaszter neve	Klaszter méret (%)
	Reumatológia, egészségügyi gazdaságtan	2.3	Egészségügyi gazdaságtan	2.0	Finika és pénzügy	1.8	Társadalom egészségügy, demográfia	1.8	Könyvelés	1.8
	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)
1	ANN RHEUM DIS	242	Eur J Health Econ	29	PHYSICA A	92	INT J OBESITY	117	ACCOUNTING J INT	91
2	ARTHRITIS RHEUM	157	NA	29	PHYS REV E	64	J HEALTH ECON	65	ACCOUNTING ACCOUNTING	39
3	RHEUMATOLOGY	113	LANCET	28	ACTA PHYS POL B	58	PEDIATRICS	52	AUDITING	26
4	EUR J HEALTH ECON	102	ANN RHEUM DIS	26	PHYS REV LETT	28	DEMOGRAPHY	39	AUST ACCOUNT	26
5	PHARMACOECONOMICS	90	Magy Onkol	24	J FINANC	24	HOUSEHOLD	39	CRIT PERSPECT	26
Társadalmi távolságmátrix	Egészségügyi gazdaságtan, Ökológiai gazdaságtan	2.7	Egészségügyi gazdaságtan	1.8	Pénzügyi menedzsment	1.8	Egészségügyi gazdaságtan	1.8	Egészségügyi gazdaságtan	1.7
	EUROPEAN JOURNAL OF HEALTH ECONOMICS	19	ACTA OECONOMICA	2	JOURNAL OF INTERNATIONAL FINANCIAL MANAGEMENT & ACCOUNTING	1	VALUE IN HEALTH	1	EUROPEAN JOURNAL OF HEALTH ECONOMICS	3
1	ECOLOGICAL ECONOMICS	1	J HEALTH	1						
Tiszta kognitív távolságmátrix	Reumatológia, egészségügyi gazdaságtan	12.2	Pénzügy, Ökonometria	5.7	Közgazdaságtan, ökonometria	4.3	Kutatáspolitikai, regionális gazdaságtan	4.0	Pénzügyi és politikai gazdaságtan	3.8
1	ANN RHEUM DIS	406	J FINANC	256	AM ECON REV	100	RES POLICY	74	AM ECON REV	89
2	ARTHRITIS RHEUM	192	AM ECON REV	100	Q J ECON	57	J POLIT ECON	36	J MONETARY	72
3	EUR J HEALTH ECON	182	ECONOMETRICA	80	J INT ECON	52	REG STUD	31	J POLIT ECON	71
4	RHEUMATOLOGY	179	PHYSICA A	76	ECON LIT	49	J ECON GEOG	23	Q J ECON	45
5	PHARMACOECONOMICS	141	ENERG ECON	69	ECONOMETRICA	47	PAP REG SCI	18	J ECON HIST	38

17. táblázat: Az öt legnagyobb szerzői klaszter öt leggyakoribb folyóirata és ez alapján a klaszter tématerület-elnevezései mindhárom távolságtípus esetében a közgazdaságtan területén

	Klaszter neve	Klaszter méret (%)	Klaszter neve	Klaszter méret (%)	Klaszter neve	Klaszter méret (%)	Klaszter neve	Klaszter méret (%)	Klaszter neve	Klaszter méret (%)
	Astronológia, geológia	3.6	Negyeddöntéskutatás, paleogeográfia	2.4	Klimatológia, biogeográfia, talajtan	2.1	Meteorológia, hidrologia, hidrometeorológia	1.9	Agártudomány és talajtan	1.7
	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)
Teljes kognitív távolságmátrix	1. ACTA ASTRONAUT	805	QUATERNARY SCI REV	1459	GLOBAL CHANGE BIOL	831	JATMOS OCEAN	402	ENVIRON	306
2	ASTROBIOLOGY	541	QUATERN INT	678	BIOGEOSCIENCES	482	IEEE T GEOSCI	277	GLOBAL CHANGE	139
3	THESES U INNSBRUCK I	225	PALAEODOL	608	SOIL BIOL BIOCHEM	289	NAT HAZARD	186	AGROECOSYS	120
4	33 INT CENV SYST	159	QUATERNARY SCI	402	NATURE	197	JAPPL METEOROL	146	SOIL USE MANAGE	99
5	GEOLOGY	150	PALAEOLINOL	340	ECOSYSTEMS	160	CLIM	146	METEOROL	68
Társadalmi távolságmátrix	Astronológia, távérsekeles	3.6	Biogeográfia, távérsekeles és	2.1	Hidrologia	2.0	Hidrologia	1.4	Biogeográfia	1.4
1	ASTROBIOLOGY	5	BIOGEOSCIENCES	3	NATURAL HAZARDS AND EARTH SYSTEM	2	HYDROLOGY AND EARTH SYSTEM	1	BIOGEOSCIENCES	2
2	IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND		GLOBAL ECOLOGY AND		HYDROLOGY AND EARTH SYSTEM					
3			REMOTE SENSING OF ENVIRONMENT		JOURNAL OF HYDROLOGY					
Tiszta kognitív távolságmátrix	Negyeddöntéskutatás, paleogeográfia	6.1	Negyeddöntéskutatás, paleogeográfia	5.4	Paleogeográfia, tektonika, geológia	3.2	Hidrologia, meteorológia	3.1	Klimaváltozás, biogeográfia	3.0
1	QUATERN INT	1550	QUATERNARY SCI REV	1729	PALAEODOL	260	HYDROL PROCESS	448	BIOL	817
2	QUATERNARY SCI REV	1011	QUATERN INT	929	ACTA GEOL HUNG	219	TECH	410	BIOGEOSCIENCES	411
3	QUATERNARY RES	503	PALAEODOL	661	TECTONOPHYSICS	155	REMOTE	283	SOIL BIOL BIOCHEM	282
4	PALAEODOL	471	QUATERNARY RES	488	LITHOS	139	DYST IC	281	NATURE	232
5	RADIAT MEAS	392	QUATERNARY SCI	459	EARTH PLANET SC LETT	123	EARTH SYS	220	ECOSYSTEMS	160

18. táblázat: Az öt legnagyobb szerzői klaszter öt leggyakoribb folyóirata és ez alapján a klaszter tématerület-elnevezései mindhárom távolságtípus esetében a természetföldrajz területén

A klaszterek jellemzése a kialakulásukat meghatározó változó (publikáció vagy hivatkozás) klaszterenkénti gyakorisága alapján történt. A 17. és 18. táblázat mindhárom távolság mentén a legnagyobb elemszámmal rendelkező öt klasztert mutatja. A klaszterek méreténél az adott klaszter elemszámának és az összes szerző számának százalékos aránya látható. Ebből is látható, hogy nagyon sok, kis elemszámú klaszter jött létre, a legnagyobb klaszterek is csupán az összes szerző 2–12%-át foglalták magukba. A 17. és 18. táblázat tartalmazza még a társadalmi távolság esetén azt az öt folyóiratot, melyben

az adott klaszterhez tartozó szerzők a leggyakrabban publikáltak (a folyóirat neve mellett az ott megjelent publikációk száma szerepel). A teljes és tiszta kognitív távolságmátrixokból kialakított klasztereknél az öt leggyakrabban hivatkozott folyóirat (és mellettük a hivatkozások száma) található.

A klaszterben található kumulált tíz leggyakoribb folyóirat alapján neveztem el a klasztereket, ehhez a Journal Citation Report-ban (JCR) található folyóiratok WCat besorolását is figyelembe vettem, igaz, a táblázatban csak az első öt folyóirat neve került feltüntetésre (17. és 18. táblázat). A közgazdaságtan területén a gyakori egészségügyi tartalom a WoS tématerületekre vonatkozó torzító hatása miatt volt jelen. Látható, hogy még a legnagyobb klaszterek is igen kis méretűek voltak, bár a tiszta kognitív mátrix valamivel nagyobb klasztereket eredményezett. A társadalmi távolságmátrix esetében jóval kevesebb folyóirat szerepelt egy-egy klaszterben és azok előfordulási gyakorisága is alacsony volt. Ennek oka az volt, hogy a vizsgált időablak (öt év) alatt sokkal kevesebb tényleges társszerzői kapcsolat jöhetett létre, mint a kognitív mátrixokban, a hivatkozásokon keresztül.

Összességében elmondhatjuk, hogy a létrejött klaszterek között nagy volt az átfedés a tématerületre vonatkozóan. Ugyanazon folyóiratok gyakran több klaszterben is megjelentek, igaz, eltérő gyakorisági értékekkel. A társadalmi, a teljes és tiszta kognitív távolság esetében tapasztalt hasonló tématerületek a kutatók hasonló tudásbázisára vezethetők vissza, hiszen a tudományos együttműködések létrejöttéhez kiemelten fontos a hasonló tudásbázis megléte.

7.2. Távolságtípusok összehasonlítása hálózatelemzési módszerekkel

A létrehozott hasonlósági mátrixok mindegyike egy-egy hálózatként is felfogható. Ez esetben a csúcsok a szerzők, az élek a hasonlósági mátrix celláiban található, nem 0-val egyenlő értékek. A két szereplő közötti kapcsolat kifejezi, hogy mennyire hasonlítanak egymáshoz, azaz mennyire vannak egymáshoz közel a társadalmi vagy a kognitív térben. A létrejött hálózatok mindegyike súlyozott és irányítatlan, hurkok és többszörös élek nem találhatók bennük. Súlyozott, mert a kapcsolatok erőssége a hasonlóság mértékével van megadva és irányítatlan, mivel a szereplők közötti kapcsolat kölcsönös. A hálózatok csúcsai a szerzők, így ezek száma egy tudományterületen belül azonos mindhárom hálózatban. Az élek száma a szerzők között definiált kapcsolat alapján változik. A társszerzői hálózat esetében két szerző között akkor van él, ha közösen publikáltak egy cikket, így ez a hálózat valós kapcsolatokon alapszik. A teljes kognitív hálózatban két szerző között akkor van kapcsolat, ha ugyanarra a cikkre hivatkoztak mindketten, ezt nevezte Yan és Ding (2012) találó elnevezéssel konstruált kapcsolatnak.

A következő alfejezetekben különböző hálózattudományi módszerekkel hasonlítom össze a teljes kognitív, a tiszta kognitív és a társadalmi távolságot leíró hálózatokat.

7.2.1. A társadalmi és a kognitív hálózatok jellemzése

A három hálózat fő paramétereit tartalmazza a 19. táblázat. Látható, hogy a kognitív hálózatok jóval több élt tartalmaznak és hálózati sűrűségük is magasabb, mint a valós kapcsolatokat tartalmazó társszerzőségi hálóké. Ugyanakkor mindezzel együtt a hálózati sűrűségértékek mindegyik esetben egy viszonylag ritka hálót mutatnak. A természetföldrajz területén meghatározott hálózatok nem csak nagyobbak, de az elméletileg lehetséges kapcsolatok nagyobb százalékát is tartalmazzák, mint a közgazdaságtan területén létrehozott hálózatok.

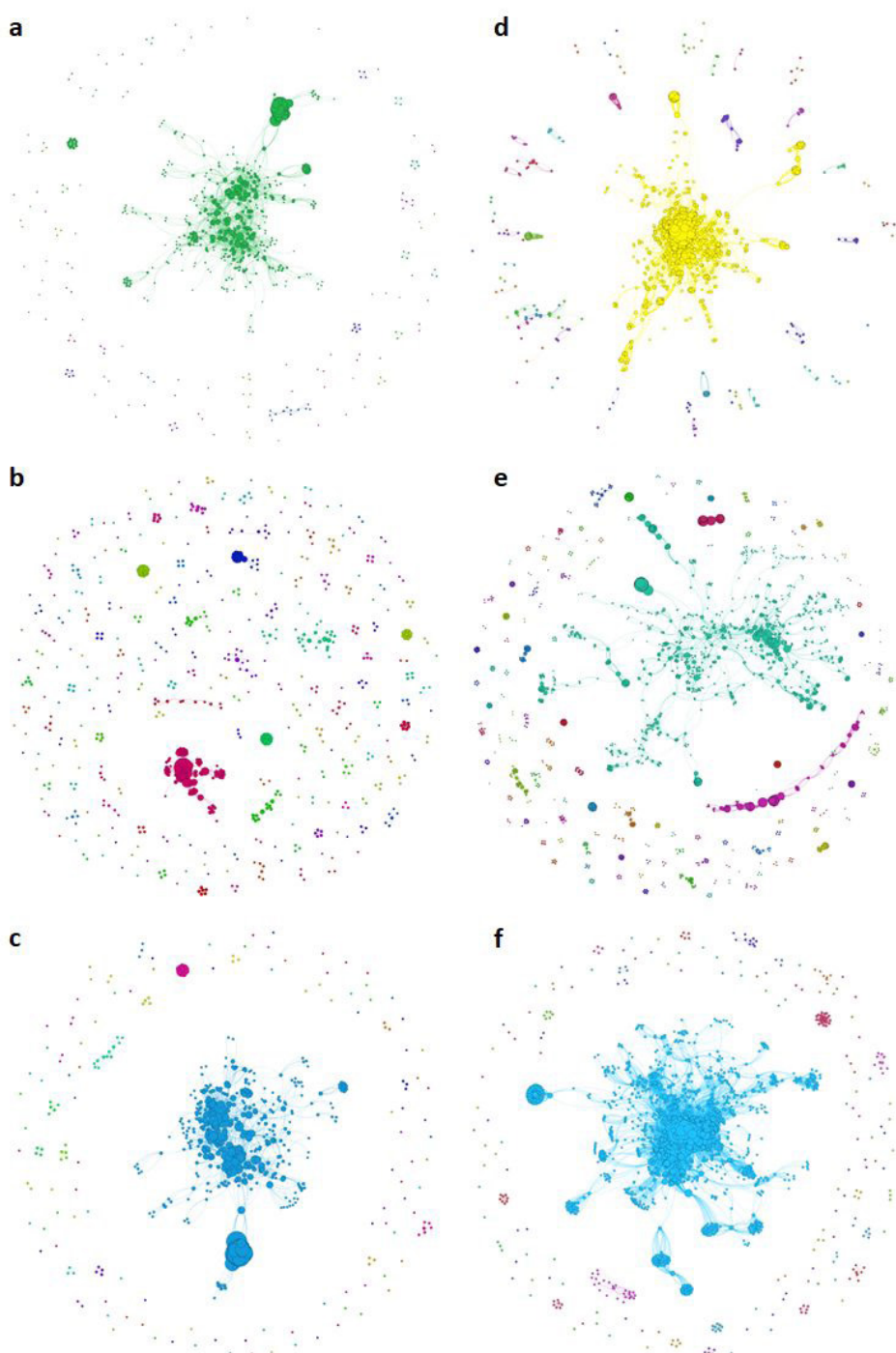
		Közgazdaságtan	Természetföldrajz
Teljes kognitív hálózat	<i>csúcsok száma</i>	704	2294
	<i>élek száma</i>	5145	67282
	<i>hálózati sűrűség</i>	2,08%	2,56%
	<i>óriáskomponens aránya</i>	74%	88%
	<i>globális klaszterezettség</i>	0,733	0,539
	<i>átlagos úthossz*</i>	3,9487	3,5495
Társadalmi hálózat	<i>csúcsok száma</i>	704	2294
	<i>élek száma</i>	1347	16351
	<i>hálózati sűrűség</i>	0,54%	0,62%
	<i>óriáskomponens aránya</i>	13%	52%
	<i>globális klaszterezettség</i>	0,682	0,927
	<i>átlagos úthossz*</i>	2,3323	6,3878
Tiszta kognitív hálózat	<i>csúcsok száma</i>	704	2294
	<i>élek száma</i>	4874	61157
	<i>hálózati sűrűség</i>	1,97%	2,33%
	<i>óriáskomponens aránya</i>	74%	88%
	<i>globális klaszterezettség</i>	0,699	0,493
	<i>átlagos úthossz*</i>	3,9503	3,5513

* a hálózat főkomponensére vonatkozó érték

19. táblázat: Hálózatok fő paramétere

A teljes és a tiszta kognitív hálózat mindkét tudományterületen hasonló képet mutat a hálózat összefüggőségét tekintve (19. táblázat): a hálózati komponensek közül a legnagyobb óriáskomponensnek nevezhető, tekintve, hogy a csúcsok többségét (közgazdaságtan: 74%, természetföldrajz: 88%) magába foglalja. A hálózat fennmaradó csúcsai sok apró komponensbe tömörülnek.

A 22. *ábra* a társadalmi, teljes és tiszta kognitív hálózatokat mutatja a két tudományterületen a hálózat komponensei alapján színezve. A hálózatok csúcsainak mérete a csúcsok fokszámával arányos. A hálózatok közepén találhatók a létrejött óriáskomponensek. Az ábrán jól látható, hogy a kognitív hálózatokhoz képest a társz-szerzői hálózatoknál jóval kisebb a legnagyobb összefüggő komponens.



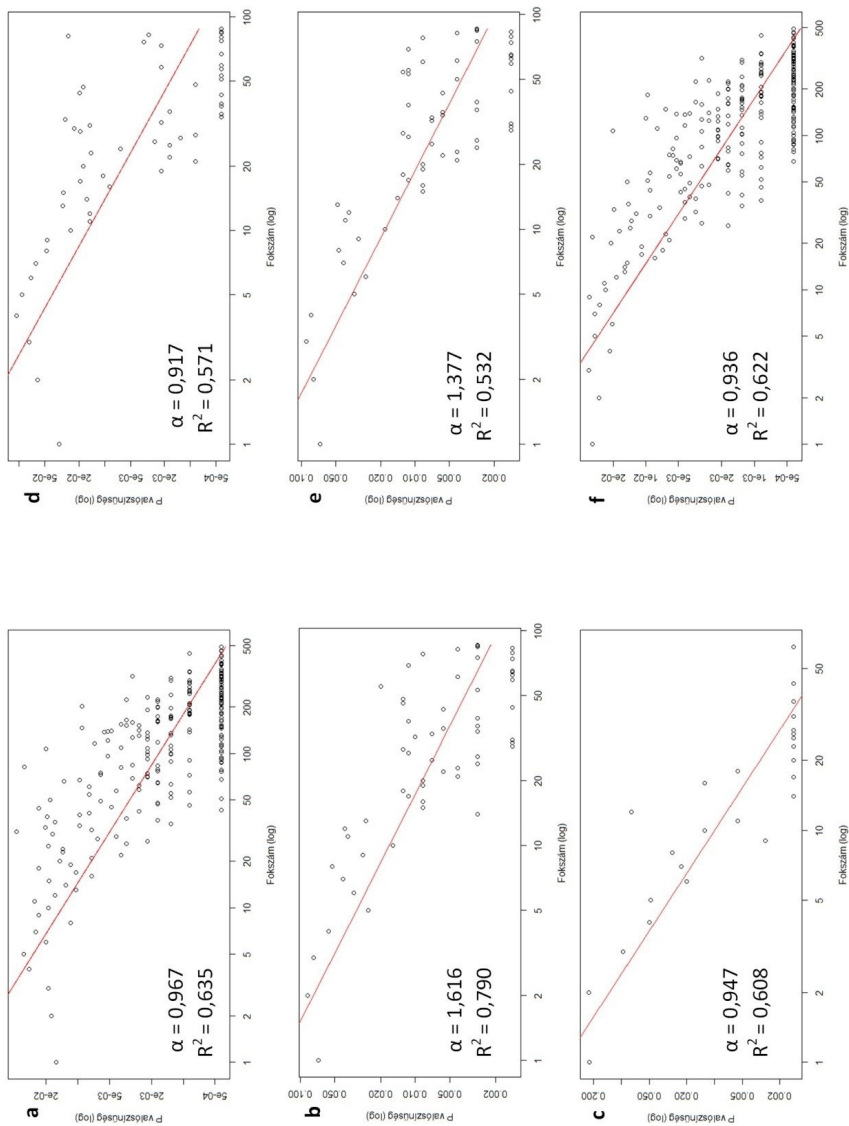
22. ábra: Teljes kognitív (a,d), társadalmi (b,e) és tiszta kognitív (c,f) hálózatok komponensei a közgazdaságtan (a,b,c) és a természetföldrajz (d,e,f) területén

Megjegyzés: a csúcsok mérete a fokszámukkal arányos, a színek a különböző komponenseket jelölik.

A közgazdaságtan területén vizsgált társszerzői hálózatban nem is jelent meg az óriáskomponens, azaz a legnagyobb összefüggő alhálózat nem foglalta magában a csúcok többségét, csupán 13%-ukat. Ez a hálózat nagyon töredezett volt, kisebb csoportok alkották. A két tudományterületen megjelenő kisebb főkomponens arra vezethető vissza, hogy a társszerzői hálózatok eleve ritkábbnak bizonyultak a vizsgált mintán: ez nem meglepő annak fényében, hogy a publikációk száma és a hozzájuk felhasznált szakirodalmi tételek száma nagyságrendi eltérést mutat az utóbbiak javára. Az pedig, hogy a közgazdaságtan területén kirajzolódó társszerzői hálózatból hiányzik az óriáskomponens, a tudományág publikációs gyakorlatát tükrözi: a hálózati ábra (22. ábra b) tanúsága szerint több, viszonylag nagyobb részgráf is kialakul, melyek között nincsen összeköttetés. Ez arra utalhat, hogy az egyes kutatócsoportok, iskolák között kevés az átjárhatóság, akkor is, ha hasonló munkákra hivatkoznak – ezt a két közgazdaságtani kognitív hálózat óriáskomponensének nagy mérete bizonyítja (22. ábra a, c). Ugyanakkor azt is meg kell jegyezni, hogy a megállapítás a vizsgált öt éves időintervallumra vonatkozik – mely egyébként a bibliometriai adatokat felhasználó tanulmányokban gyakran alkalmazott időablak (Boyack és Klavans, 2010) – így természetesen elképzelhető, hogy bizonyos múltbeli együttműködésekre nem derített fényt a vizsgálat. Mindezek ellenére a tendenciát jól jelzi, hogy a többi hálózat rendelkezett óriáskomponenssel.

A globális klaszterezettség mindegyik hálózat esetében viszonylag magas értékeket mutatott. A hálózatok közül a természetföldrajz területén vizsgált társszerzői hálózat volt a legklaszterezettebb, azaz itt voltak a leggyakoribbak a hálózaton belül a háromszögek. Newman (2001a) több tudományág társszerzői hálózatait vizsgálva a globális klaszterezettségre 0,4 és 0,7 közötti értékeket mutatott ki (kivéve az orvosbiológia területén meghatározott társszerzői hálót, melynek klaszterezettsége csupán 0,066 volt). Az általam vizsgált társszerzői hálózatok ehhez hasonló, illetve ennél magasabb klaszterezettséget mutattak (19. táblázat).

Megvizsgáltam a hálózatok fokszámoszlását is. A hálózatok fokszámoszlásai közel hatványfüggvény-eloszlást mutattak (23. ábra), mivel az α kitevő értéke az 1 értéket közelítette (Barabási, 2016). A hatványfüggvények illeszkedését az R^2 értékek is megfelelően alátámasztották.



23. ábra: Teljes kognitív (a,d), társadalmi (b,e) és tiszta kognitív (c,f) hálózatok fokszámeloszlása a közgazdaságtan (a,b,c) és a természetföldrajz (d,e,f) területén

R^2 a hatványfüggvényhez való illeszkedés mértéke, α a hatványkitevő

A 19. táblázat a főkomponensekre (a legnagyobb összefüggő alhálózatra) vonatkozóan mutatja az átlagos úthosszt, mely a csúcspárok között meghatározott legrövidebb utak átlaga. Nem összefüggő hálózat esetén az átlagos úthosszt a főkomponensre érdemes meghatározni, hiszen ha az egyes részgráfok között nincs összeköttetés, akkor a két nem összefüggő alhálózat csúcsai között a távolság ∞ értéket vesz fel. A vizsgálat során én is az óriáskomponensekre vonatkozóan határoztam meg a hálózatok átlagos úthosszát. A számításokhoz használt R szoftver igraph csomagját az egész hálózatra futtatva megadható, hogy a hálózat összefüggő-e. Amennyiben nem, szintén a legnagyobb komponensre történik az átlagos úthossz meghatározása. Félrevezető eredményhez vezethet, ha ezzel a beállítással nem élünk: ilyenkor a nem összefüggő hálózat részhálózataiban található csúcsok között ∞ távolság-érték helyett 0 értékkel számol.

A teljes és tiszta kognitív hálózatok a 19. táblázat szerint mind a természetföldrajz, mind a közgazdaságtan területén egymással szinte azonos átlagosúthossz-értéket mutattak. Ha a két tudományterület átlagos úthosszát hasonlítjuk össze, a közgazdaságtan területén ténylegesen és arányaiban is magasabb átlagos hálózati távolságot (3,95) találunk, mint a természetföldrajzi hálózatban (3,55). Ez a két hálózat eltérő méretéből, sűrűség- és átlagosfokszám-értékéből következik. Az ok a két tudományterület eltérő hivatkozási szokásaiban keresendő: a közös hivatkozások előfordulása a természetföldrajz esetén gyakoribbnak mutatkozott, ez pedig a kognitív hálózatokban több élként jelent meg.

A teljes és tiszta kognitív hálózatok nagyobb sűrűségmutatókkal rendelkeztek, és az óriáskomponens is a hálózat csúcsainak a többségét magába foglalta. A jóval ritkább társszerzői hálózatok átlagos úthosszánál emiatt magasabb értékeket feltételeztem: a természetföldrajzi adatokon tapasztalt érték ezt megerősítette, itt 6,4 volt az átlagos úthossz. A közgazdaságtan területén azonban igen alacsony (2,3) átlagos úthosszt találtam. Ez azzal magyarázható, hogy a közgazdaságtan területén a társszerzői hálózat több kisebb komponensből állt, és ezek közül a főkomponens a szerzők csupán 13%-át foglalta magába. Ebben a kicsi, csak 88 csúcsból álló alhálózatban nem meglepő a kapott 2,3 lépés átlagosúthossz-érték. Összességében a hálózatok méretéhez képest elég magas átlagosúthossz-értékeket találtam.

Ezt követően az egyes szerzők és a többi szerző közötti legrövidebb utak átlagát határoztam meg, és Pearson-féle korreláció segítségével hasonlítottam össze az egyes hálózatok értékeit (20. és 21. táblázat). Az eredmények azt mutatták, hogy a két kognitív hálózat egymáshoz nagyon hasonló, míg a társszerzői hálózat erős kapcsolatot mutatott mindkét kognitív hálózattal.

	Teljes kognitív hálózat	Társadalmi hálózat	Tiszta kognitív hálózat
Teljes kognitív hálózat	1	0,74040	0,99998
Társadalmi hálózat	0,74040	1	0,74038
Tiszta kognitív hálózat	0,99998	0,74038	1

20. táblázat: Szerzők közötti átlagos hálózati távolságok Pearson-korrelációja, közgazdaságtan

	Teljes kognitív hálózat	Társadalmi hálózat	Tiszta kognitív hálózat
Teljes kognitív hálózat	1	0,64236	0,99998
Társadalmi hálózat	0,64236	1	0,64229
Tiszta kognitív hálózat	0,99998	0,64229	1

21. táblázat: Szerzők közötti átlagos hálózati távolságok Pearson-korrelációja, természetföldrajz

A teljes és a tiszta kognitív hálózat a hálózat alapparamétereit tekintve (19. táblázat) nagyfokú hasonlóságot mutatott. Ezek a paraméterek a hálózat szerkezetéről is információt nyújtanak, így abban is feltételezhetően hasonló a két kognitív hálózat.

7.2.2. A társadalmi, a teljes és tiszta kognitív hálózatok összehasonlítása QAP korreláció segítségével

A létrehozott hálózatok közötti összefüggéseket az egész hálózatokra vonatkozó Quadratic Assignment Procedure (QAP) korreláció segítségével is vizsgáltam. A módszer lehetővé teszi hálózatok, hasonlósági és távolságmátrixok összehasonlítását (Dietz, 1983; Dow és Cheverud, 1985), valamint hipotézisek tesztelését (Krackardt, 1987). Széles körben elterjedt, több tudományterületen is alkalmazzák, mint például biológia (Legendre, 2000), pszichológia (Hubert, 1986), tudománymetria (McCain et al., 2005), de leggyakrabban a szociológia területén alkalmazzák a kapcsolathálózat-elemzés (*Social Network Analysis, SNA*) egyik módszereként (Cross et al., 2001; Lee et al., 2003; Borgatti és Cross, 2003; Burris, 2005; Bell, 2005; Robins, 2013; Everett és Borgatti, 2014). A módszer előnye, hogy értelmezése azonos a Pearson-féle korrelációval és viszonylag kis számítási igényű.

A hagyományos korreláció a különböző jelenségek, adatsorok közötti összefüggések vizsgálatára gyakran alkalmazott matematikai statisztikai eljárás. A hálózatok vizsgálatakor rendelkezünk adatokkal, amelyek az egyes hálózati szereplőkre, azaz a csúcsokra vonatkoznak: ezeken az adatokon végezhetünk hagyományos korrelációs számításokat. Azonban a hálózatoknál rendelkezünk a szerkezetre vonatkozó adatokkal is: ezek az élek, amelyek meghatározzák, hogy mely szereplők között léteznek kapcsolatok, diadikus viszonyok.

Az összehasonlításra kerülő diadikus viszonyok nem függetlenek egymástól, és ez sokáig gátat vetett a megbízható statisztikai tesztek kialakításának (Proctor, 1969; Laumann és Pappi, 1976). Emiatt nem állapítható meg a t- vagy F-próbával a hagyományos lineáris korreláció során kapott eredmények szignifikanciája. Ennek orvoslására ún. permutációs tesztet alkalmaznak, melynek neve Quadratic Assignment Procedure. Tehát a QAP megközelítés egy speciális permutációs teszt típus, mely megtartja a diadikus adatstruktúrákat a permutáció alatt. A statisztikus Mantel dolgozta ki 1967-ben, ő is alkalmazta először a járványügyi kérdéssel foglalkozó tanulmányaiban, melyekben azt vizsgálta, hogy vajon a betegségek terjedése földrajzilag klaszterezett-e. A permutációs eljárást róla nevezték el Mantel-tesztnek (Mantel 1967; Hubert és Schultz 1976; Krackhardt, 1988). A *quadratic assignment* elnevezés Hubert (1986) nevéhez kötődik.

A QAP korreláció azt mutatja meg, hogy ha egy adott hálózatban két szereplő között létezik kapcsolat, akkor mekkora a valószínűsége annak, hogy egy másik hálózatban is lesz kapcsolat ugyanezen két szereplő között. Ha a korreláció értéke 1, akkor a két hálózatban ugyanazon szereplők között van kapcsolat, azaz a két hálózat megegyezik egymással. Amennyiben a korreláció értéke -1, akkor az egyik hálózatban, ha két szereplő között van kapcsolat, akkor a másik hálózatban ugyanezen szereplők között nincs kapcsolat és fordítva (Krackhardt, 1987).

A QAP korreláció alkalmazásához szükséges, hogy az összehasonlítandó hálózatok szereplői megegyezzenek egymással és az összehasonlítás alapja két négyzetes mátrix legyen. Ez a feltétel jelen kutatás során teljesül, hiszen mindhárom hálózat esetén ugyanazok a szerzők a csúcsok. A permutációs teszt lényege, hogy a két hálózat között a korrelációs számítás akár több ezer-szer-tízezerszer megtörténik. A teszt a következőképpen zajlik: először a két mátrix között kiszámoljuk a korrelációt, majd az egyik mátrix sorait és oszlopait azonos módon véletlenszerűen összekeverjük úgy, hogy a hálózaton belül a diadikus viszonyok azonosak maradjanak. Ezt követően a permutált és az eredeti mátrix között megismételjük a korreláció számítását. Az összekeverés és korrelációs számítás lépései sokszor ismétlődnek, a kapott korrelációs együtthatókat mindig az eredeti értékkel vetjük össze. Amennyiben az

ismétlések kevesebb mint 5%-ában kapunk eredményül az eredeti korrelációnál magasabb értéket, a két hálózat közötti kapcsolat szignifikáns.

A módszer előnye a már korábban említett viszonylag kicsi számításigény, emiatt hipotézisek gyors tesztelésére és az eredmények könnyű értelmezésére jól alkalmazható. Hátránya, hogy csak egész hálózatok összehasonlítására van lehetőség és egy hálózat időbeli változásának becslésére nem igazán alkalmas. Jelen vizsgálatban azonban nagyon jól alkalmazható, mivel gyorsan ad megbízható eredményt a hasonlósági mátrixok egymáshoz való viszonyáról, és megmutatja, mennyire hasonlítanak egymáshoz a kapcsolatok az egyes hálózatokban.

A QAP korrelációt ezer ismétlésszámmal futtattam le, minden esetben a 0,05-ös szinten szignifikáns eredményt kaptam. A 22. és 23. táblázatban láthatók a QAP korrelációs együtthatók a közgazdaságtan és a természetföldrajz területén. Mindkét esetben a teljes kognitív hálózat a társszerzői hálózattal erős kapcsolatot mutatott, viszont a tiszta kognitív hálózatot mindkét másik hálózattal összevetve gyenge korrelációt találtam. A társszerzői kapcsolatok kiszűrését követően láthatóan a teljes kognitív és a tiszta kognitív hálózat között gyenge kapcsolat mutatkozik. Az eredmények alapján elmondható, hogy a társszerzői hálózat és a teljes kognitív hálózat között a legerősebb az összefüggés, azaz a társszerzőség a teljes kognitív hálózatban meghatározóbb faktorként jelenik meg, mint a tiszta kognitív hálózat. Ez arra enged következtetni, hogy a teljes kognitív hálózatban a közös hivatkozások jellemzően a társszerzőségekből következnek. Ebből a szempontból a két tudományterület hasonló képet mutat (22. és 23. táblázat). A természetföldrajz esetében a társadalmi komponens dominánsabb, de a különbség a két terület között elenyésző.

	Teljes kognitív hálózat	Társadalmi hálózat	Tiszta kognitív hálózat
Teljes kognitív hálózat	1	0,9872	0,3170
Társadalmi hálózat	0,9872	1	0,2375
Tiszta kognitív hálózat	0,3170	0,2375	1

22. táblázat: QAP korreláció eredménye a közgazdaságtan területén

	Teljes kognitív hálózat	Társadalmi hálózat	Tiszta kognitív hálózat
Teljes kognitív hálózat	1	0,9896	0,3280
Társadalmi hálózat	0,9896	1	0,2564
Tiszta kognitív hálózat	0,3280	0,2564	1

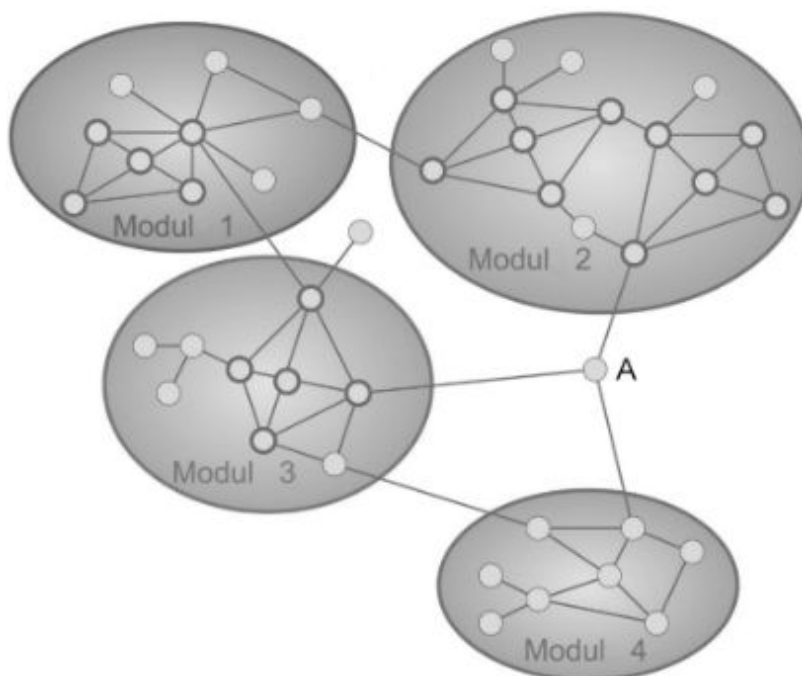
23. táblázat: QAP korreláció eredménye a természetföldrajz területén

7.2.3. Szerzői csoportok által kijelölt tématerületek az egyes távolságmátrixokban hálózati modularitással

A hálózatok kisebb alhálózatokra, összefüggő komponensekre bonthatók. A hálózatok kisebb csoportokra történő bontása régóta jelen levő jogos igény. A hálózaton belüli csoportokat moduloknak is nevezik, és attól függően, hogy mi alapján történik a hálózati csoportok meghatározása, számos csoportképző algoritmus létezik. Newman (2006) a csoportkereső eljárások két fő típusát különítette el. Az egyiket gráffelosztásnak nevezte, ekkor előre meghatározott számú és méretű csoportokra kell bontani a hálózatot. Ezt főként a számítástechnikában alkalmazzák, például a processzormagok kapacitás-elosztásának megtervezésekor. A másik fő típuscsoport a különböző csoportkeresési (*community detection*) eljárásokat foglalja magában, ekkor a hálózat szerkezete határozza meg a létrejövő csoportok számát és méretét, ezt főképp a szociológia, a fizika, az alkalmazott matematika és a biológia területén alkalmazzák. A meghatározott hálózati csoportok vagy modulok csúcsai egymás között szoros kapcsolódást mutatnak mind szerkezetileg, mind funkcionálisan, azonban a hálózat többi eleme felé kevesebb kapcsolattal rendelkeznek, így azoktól viszonylag elkülönülnek (Csermely, 2005). A 24. ábrán például látható, hogy az egyes modulokhoz tartozó csúcsok a modul többi csúcsával vannak inkább összeköttetésben, míg az ahhoz nem tartozókkal kevés kapcsolaton keresztül kötődnek, vagy azoktól külön állnak.

A létező számos csoportkeresési eljárásból a disszertációban én az ún. fast greedy modularitás eljárást alkalmaztam (Clauset et al., 2004). Ez egy hierarchikus csoportkeresési algoritmus, mely a greedy optimalizációt használja és nagy hálózatokon is gyorsan meghatározza a csoportokat. Mivel a valóságot leíró hálózatok általában nagyok és ritkák, valamint jellemző rájuk a hierarchikus felépítés, ezeken sok szokásos és régebbi algoritmus csak nagyon lassan vagy nem is ad eredményt. A szerzők (Clauset et al., 2004) egy gyors és hatékony, alulról fölfelé építkező eljárást dolgoztak ki ennek orvoslására.

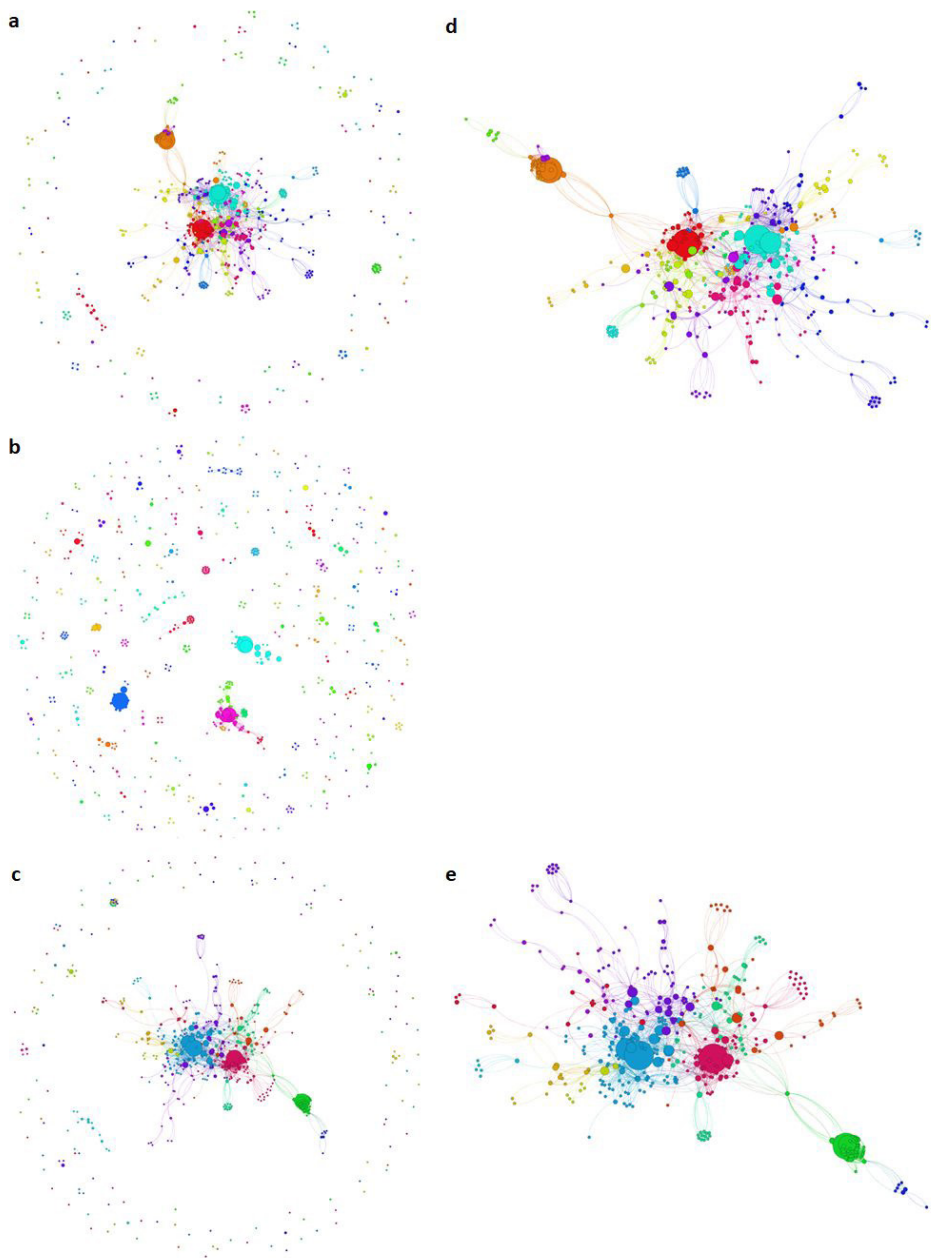
Korábban Girvan és Newman (2002) egy, a közteségcentralitáson alapuló eljárást is meghatározott, mely a hálózat csoportjait azok perifériájának azonosításán keresztül határozza meg. Így a módszer a modulok határait keresi meg, hiszen a magas közteségcentralitással rendelkező csúcsok tulajdonképpen egyik csoporthoz sem tartoznak igazán, viszont megteremtik a kapcsolatot a hálózat moduljai között. Ilyen csúcs a 24. ábra A csúcsa. A vizsgálatomban alkalmaztam ezt a csoportkeresési algoritmust is, azonban a természetföldrajz területén meghatározott hálózaton ez nagyon lassúnak bizonyult, így emiatt is esett a választásom a fast greedy algoritmusra.



24. ábra: A modularitás sematikus példája irányítatlan hálózatokon

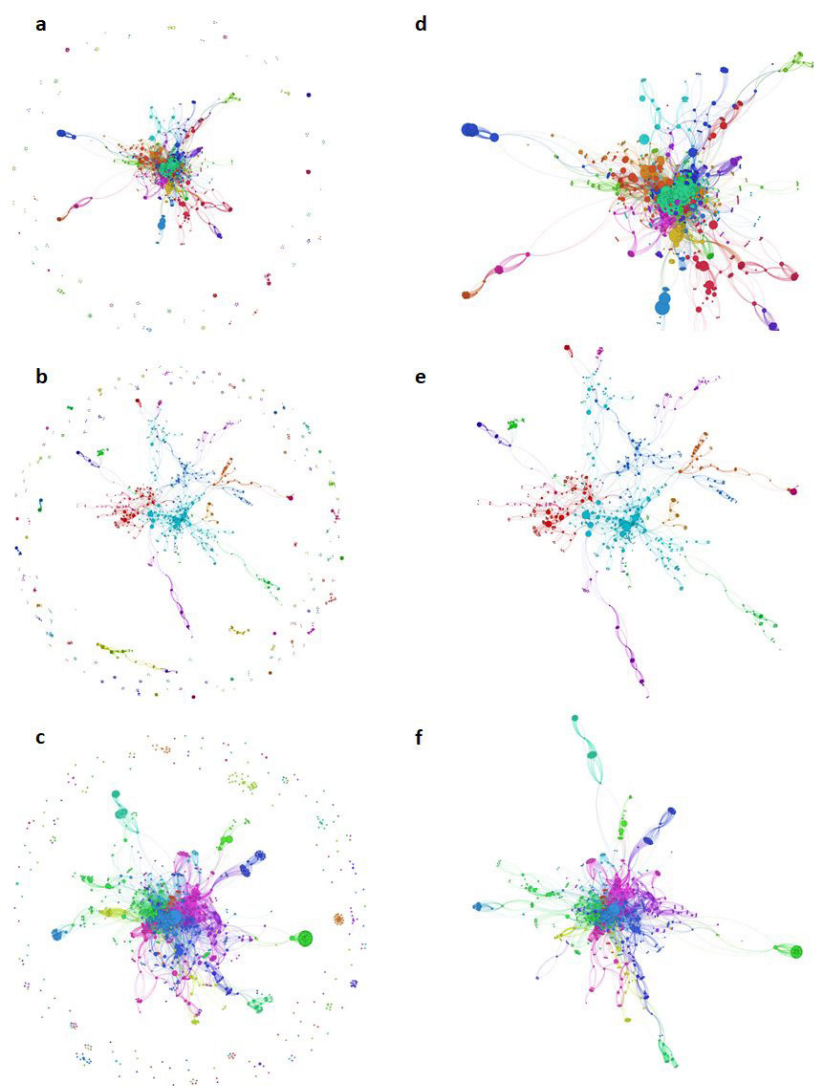
Forrás: Rollié, Mangold, Sundmacher 2012, 11. o. alapján saját szerkesztés

A 25. ábra a közgazdaságtan területén meghatározott teljes kognitív (25. ábra a, d) társadalmi (25. ábra b) és tiszta kognitív (25. ábra c, e) hálózatot (25. ábra bal oldala) és óriáskomponensét (25. ábra jobb oldala) mutatja (a társadalmi hálózatnál hiányzott az óriáskomponens). Az ábrán a különböző színnek a fast greedy algoritmus alapján meghatározott egyes modulokat jelölik. A csúcsok mérete az adott szerző publikációinak a számával arányos, azaz minél több cikket publikált egy szerző a vizsgált öt év alatt, annál nagyobb a csúcs átmérője. A legnagyobb modulok jellemzően a hálózat főkomponensében találhatók. A többi összefüggő alhálózati részben az összes szerző egy csoportba került besorolásra.



25. ábra: Teljes kognitív (a,d), társadalmi (b) és tiszta kognitív (c,e) hálózatok fast greedy algoritmus alapján létrehozott moduljai (a,b,c) és óriáskomponenseik (d,e) a közgazdaságtan területén

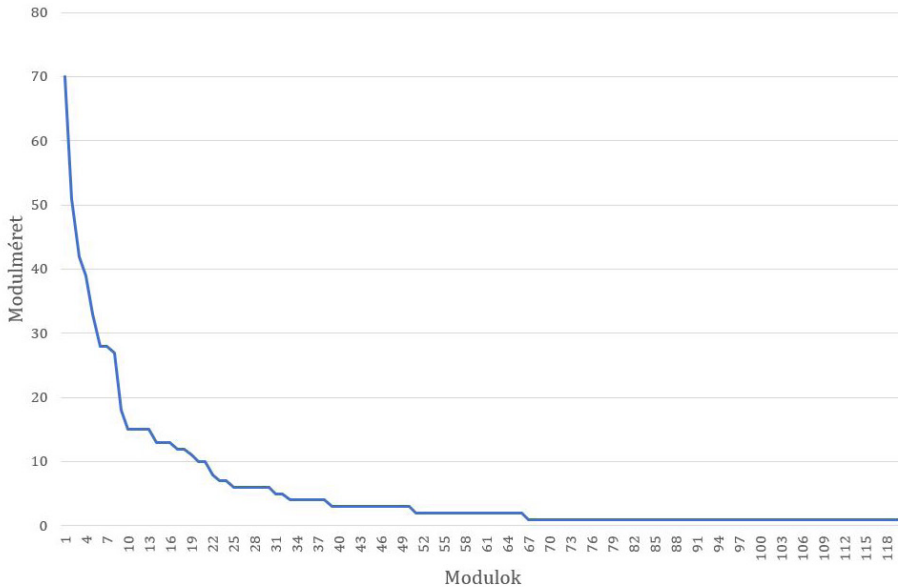
Megjegyzés: a csúcsok mérete az egyes szerzők publikációinak számával arányos; a társadalmi hálózatban nem jött létre óriáskomponens.



26. ábra: Teljes kognitív (a,d), társadalmi (b,e) és tiszta kognitív (c,f) hálózatok fast greedy algoritmus alapján létrehozott moduljai (a,b,c) és óriáskomponenseik (d,e,f) a természetföldrajz területén

Megjegyzés: a csúcsok mérete az egyes szerzők publikációinak számával arányos.

A 26. ábra a természetföldrajz területén meghatározott teljes kognitív (26. ábra a, d) társadalmi (26. ábra b, e) és tiszta kognitív (26. ábra c, f) hálózatot (26. ábra bal oldala) és óriáskomponensét (26. ábra jobb oldala) mutatja. Az ábrán a színek itt is a fast greedy algoritmus alapján létrejött modulok szerint lettek jelölve. A csúcsok mérete az adott szerző publikációinak a számával arányos. Az óriáskomponensben található itt is jellemzően a legnagyobb modulok, míg a többi hálózati komponens egy modulba került besorolásra. A két tudományterület hasonló képet mutatott.



27. ábra: Modulok méretének eloszlása a teljes kognitív hálózaton a közgazdaságtan területén

A 27. ábra a közgazdaságtan területén a teljes kognitív hálózaton meghatározott modulok méretének az eloszlását mutatja. A hierarchikus klaszteranalízis csoportjainak eloszlásához hasonló a modulok méreteloszlása. Az első, a hatodik és a nyolcadik modul méretét követően látható nagyobb törés. A vizsgálatban a hierarchikus klaszteranalízis csoportjainak jellemzéséhez hasonlóan az öt legnagyobb modult elemeztem itt is. A 24. táblázat a vizsgált két tudományterület három-három hálózatában a modulok számát és a legnagyobb öt és tíz modul egész hálózathoz viszonyított együttes arányát mutatja. Mindkét terület mindhárom hálózatában nagyon sok csoportot generált a fast greedy algoritmus, azonban ezen csoportok többsége csupán egy vagy néhány szerzőből állt.

A természetföldrajz területén a teljes kognitív hálózat legnagyobb öt csoportja a szerzők 40%-át tartalmazta, míg a tíz legnagyobb modul már a szerzők több mint felét foglalta magába. Csupán tíz egyelemű modul jött létre. A társadalmi hálózat 162 moduljából húsz egyelemű volt, a tíz legnagyobb modul pedig a szerzők közel felét tömörítette. A tiszta kognitív hálózatban volt található a legtöbb modul a természetföldrajz területén, azonban ezek közül csak 35 modul tartalmazott egynél több szerzőt, a tíz legnagyobb modul pedig összesen a szerzők háromnegyedét foglalta magába.

A közgazdaságtan területén a társadalmi hálózat rendelkezett a legtöbb csoporttal, ezek több mint a fele egyszemélyes volt, ugyanakkor csak a tíz legnagyobb modul tartalmazott tíznél több szerzőt, ezek a modulok összesen a szerzők egynegyedét tömörítették magukba. A teljes kognitív hálózat legnagyobb öt modulja a szerzők egyharmadát, míg a legnagyobb tíz a szerzők közel felét tartalmazta, az összes modul közel fele egyszemélyes volt. A tiszta kognitív hálózat csupán 20 modulja tartalmazott egynél több szerzőt. A tíz legnagyobb modul a szerzők 70%-t, míg az öt legnagyobb a szerzők több mint felét tömörítette magába. A 24. táblázat ezeket az eredményeket összegzi.

	Közgazdaságtan		Természetföldrajz	
	10 legnagyobb modul szerzőinek aránya az összes szerzőből (%)	5 legnagyobb modul szerzőinek aránya az összes szerzőből (%)	10 legnagyobb modul szerzőinek aránya az összes szerzőből (%)	5 legnagyobb modul szerzőinek aránya az összes szerzőből (%)
Teljes kognitív hálózat	48	33	57	40
Társadalmi hálózat	25	17	47	34
Tiszta kognitív hálózat	70	55	76	57

24. táblázat: A legnagyobb modulok szerzőinek aránya a teljes szerzőszámhoz viszonyítva a közgazdaságtan és a természetföldrajz területén

A teljes és a tiszta kognitív hálózat tíz legnagyobb modulja mindkét tudományterületen teljes egészében az óriáskomponens részét alkotó szerzőkből állt. A társadalmi hálózatok ettől eltérő képet mutattak: láttuk a 7.2.1. alfejezetben, hogy az óriáskomponens is jóval kisebb méretű bennük (sőt, a közgazdaságtan területén ki sem alakult), és a tíz legnagyobb modul egy része is az óriás-, illetve főkomponensen kívül található: a természetföldrajz esetén három, a közgazdaságtani hálózatban pedig hét modult is valamelyik kisebb szerzői alhálózatban találunk.

A hálózati modularitás a hierarchikus klaszteranalízishez hasonlóan csoportokat alakít ki. Ebben az esetben azonban a hálózat szerkezete alapján jönnek létre a csoportok. A kialakult szerzői csoportokat a 7.1.2. alfejezetben bemutatott módosított Rand-index segítségével hasonlítottam össze, a hierarchikus klaszteranalízis által létrejött szerzői klaszterek összehasonlításához hasonlóan. A 25. és 26. táblázat a fast greedy algoritmus alapján létrejött szerzői csoportok módosított Rand-indexszel történt összehasonlítását mutatja. Mind a két tudományterületen a teljes és a tiszta kognitív hálózatokban létrehozott szerzői csoportok bizonyultak egymáshoz hasonlóknak. A társadalmi hálózat csoportjai különböztek a kognitív hálózatok csoportjaitól. A teljes és tiszta kognitív hálózat szerzőinek hasonló csoportbeosztása az algoritmusra vezethető vissza. Bár a fast greedy algoritmus figyelembe veszi az élsúlyokat, mégis, mivel a hálózat élein halad keresztül, a hálózat szerkezete nagy befolyással bír a csoportok kialakulására. Ahogy a 7.2.1. alfejezetben a hálózatok fő paramétereit bemutató 19. táblázatban láttuk, a teljes és tiszta kognitív hálózat mind a két tudományterületen nagyon hasonló paraméterekkel rendelkezett. E két hálózat szerkezete sokkal inkább hasonlít egymásra, mint a társadalmi hálózat a kognitív hálózatokra.

	Teljes kognitív hálózat	Társadalmi hálózat	Tiszta kognitív hálózat
Teljes kognitív hálózat	1	0,0068	0,5689
Társadalmi hálózat	0,0068	1	0,0092
Tiszta kognitív hálózat	0,5689	0,0092	1

25. táblázat: A szerzők hálózati modularitás alapján történő csoportokba sorolásának hasonlósága módosított Rand-index alapján a közgazdaságtan területén

	Teljes kognitív hálózat	Társadalmi hálózat	Tiszta kognitív hálózat
Teljes kognitív hálózat	1	0,0066	0,5839
Társadalmi hálózat	0,0066	1	0,0130
Tiszta kognitív hálózat	0,5839	0,0130	1

26. táblázat: A szerzők hálózati modularitás alapján történő csoportokba sorolásának hasonlósága módosított Rand-index alapján a természetföldrajz területén

Az öt legnagyobb szerzői modult az azokban előforduló leggyakoribb folyóiratnevek alapján (27. és 28. táblázat) neveztem el, megegyezően a 7.1.2. alfejezetben a hierarchikus klaszteranalízis eredményeként létrejött csoportok elnevezési módjával. A klaszteranalízis öt legnagyobb csoportjának méretéhez viszonyítva a modularitás során meghatározott csoportok nagyobbak voltak, mivel ez utóbbi módszer a hálózat összeköttetései mentén hozza létre a csoportokat.

A teljes kognitív és a tiszta kognitív hálózat a közös hivatkozásokon alapult, és mivel ez a társszerzői hálózathoz nagyobb sűrűségű hálózatot eredményez, az egyes modulokba tartozó szerzők által egy bizonyos folyóiratban található cikkekre történt hivatkozások száma jóval magasabb gyakorisági értékeket vett fel, mint a társadalmi hálózat folyóirat-gyakoriságai. Ennek oka, hogy a társszerzői hálózathoz a folyóirat-gyakoriságok csupán az alapmintából származtak, míg a két kognitív hálózat közös hivatkozásai az alapmintán kívülre is mutathattak. A 27. és 28. táblázat társadalmi hálózatainak legnagyobb moduljai az adott modul szerzői által társszerzőségben publikált cikkek számát mutatják folyóiratonként.

A teljes és tiszta kognitív, valamint a társadalmi hálózatokban az egyes tudományterületeken első ránézésre hasonló témacsoportok jelentek meg. A témabesorolást nehezítették a multidiszciplináris (*Science*, *PLOS*, *Nature*), illetve a földrajzi régió szerinti tudományterületi folyóiratok (pl. *Central European Journal of Geosciences*).

	Hálózati csoport neve	Csoport méret (%)	Hálózati csoport neve	Csoport méret (%)	Hálózati csoport neve	Csoport méret (%)	Hálózati csoport neve	Csoport méret (%)	Hálózati csoport neve	Csoport méret (%)
Teljes kognitív hálózat	Reumatológia, gyógyvezérgazdaságtan	9,9	Ökonometria, pénzügyi gazdaság	7,2	Pénzügygazdaságtan, politikai gazdaságtan, kockázati biztosítás tudomány	6,0	Pénzügygazdaságtan, banki gazdaságtan	5,5	Játekelmélet, biheviória gazdaságtan	4,7
	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)
	1 AM RHEUM DIS	386	AM ECON REV	158	AM ECON REV	97	J FINANC	270	AM ECON REV	138
	2 ARTHRITIS RHEUM	192	J POLIT ECON	104	J MONETARY ECON	65	PHYSICA A	97	ECONOMETRICA	94
	3 RHEUMATOLOGY	169	AM J AGR ECON	86	J POLIT ECON	56	J FINANC ECON	73	Q J ECON	90
	4 EUR J HEALTH ECON	165	ECONOMETRICA	80	Q J ECON	50	ENERG ECON	69	GAME ECON BEHAV	66
	5 LANCET	126	Q J ECON	80	J COMP ECON	30	J BANK FINANC	67	J ECON THEORY	61
Társadalmi hálózat	Egészséggazdaságtan, ökológiai gazdaságtan, munkagazdaságtan	6,0	Poszt-kommunista gazdaságtan, agrárgazdaságtan	3,3	Egészséggazdaságtan	3,0	Egészséggazdaságtan	2,6	Pénzügygazdaságtan	2,0
	1 EUROPEAN JOURNAL OF HEALTH ECONOMICS	19	ACTA OECONOMICA	8	EUROPEAN JOURNAL OF HEALTH ECONOMICS	5	ACTA OECONOMICA	4	EUROPEAN JOURNAL OF FINANCE	1
	2 ECOLOGICAL ECONOMICS	1	POST-COMMUNIST ECONOMIES AGRICULTURAL ECONOMICS-ZEMEDLSKA	5	VALUE IN HEALTH	1	JOURNAL OF HEALTH ECONOMICS	1	JOURNAL OF ECONOMIC INTERACTION AND COORDINATION	1
	3 LABOUR ECONOMICS	1	EKONOMIKA	4					QUANTITATIVE FINANCE	1
	4		EKONOMICKY CASOPIS	3						
	5		JOURNAL OF AGRICULTURAL ECONOMICS	3						
Tiszta kognitív hálózat	Politikagazdaságtan, ökonometria, agrárgazdaságtan	18,5	Egészséggazdaságtan, reumatológia, gyógyvezérgazdaságtan	12,4	Pénzügy, bankgazdaságtan	10,7	Pénzügygazdaságtan, politikagazdaságtan	8,7	Ökonometria, játekelmélet, biheviória gazdaságtan	4,7
	1 AM ECON REV	242	ANN RHEUM DIS	406	J FINANC	381	AM ECON REV	110	AM ECON REV	146
	2 J POLIT ECON	169	ARTHRTIS RHEUM	192	J FINANC ECON	147	ECOL ECON	91	ECONOMETRICA	95
	3 Q J ECON	148	EUR J HEALTH ECON	182	PHYSICA A	97	J MONETARY ECON	65	Q J ECON	92
	4 ECONOMETRICA	163	RHEUMATOLOGY	179	ACCOUNTING EUROPE	91	J POLIT ECON	56	GAME ECON BEHAV	65
	5 J ECON LIT	100	PHARMACOECONOMICS	141	REV FINANC STUD	87	Q J ECON	52	J ECON THEORY	51

27. táblázat: Az öt legnagyobb szerzői modul öt leggyakoribb folyóirata és ez alapján a modulok tématerület elnevezései mindhárom távolságtípus esetében a közgazdaságtan területén

	Hálózati csoport neve	Csoport méret (%)	Hálózati csoport neve	Csoport méret (%)	Hálózati csoport neve	Csoport méret (%)	Hálózati csoport neve	Csoport méret (%)	Hálózati csoport neve	Csoport méret (%)
Teljes kognitív hálózat	Negyedidőszak, paleogeográfia.	21.0	Környezeti mikrobiológia	5.1	Számítógépes képalkotás.	5.1	Astrobiológia, környezeti	4.5	Biogeográfia, ökológia	4.4
	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)	Leggyakoribb folyóiratok	Gyakoriság (db)
	1 QUATERNARY SCI REV	2961	1 ICARUS	117	1 ENG GEOL	111	1 ACTA ASTRONAUT	808	1 GLOBAL ECOL BIOGEOG	252
	2 QUATERN INT	2567	2 APPL ENVIRON MICROB	107	2 IEEE T PATTERN ANAL	92	2 ASTROBIOLOGY	541	2 ECOL LETT	231
	3 PALAEOGEOGR PALAEOCL	1790	3 INT J SYST EVOL MICR	101	3 INT J COMPUT VISION	62	3 THESIS U INNSBRUCK I	225	3 ECOL INDIC	225
	4 J VOLCANOL GEOTH RES	1592	4 GEOMICROBIOL	92	4 INT J ROCK MECH MIN	52	4 REMOTE SENS ENVIRON	193	4 BIOGEOGR	221
Társadalmi hálózat	Negyedidőszak, paleogeográfia.	13.2	Geológia, geomorfológia	7.8	Astrobiológia, hidrogeológia, geomikrobiológia	5.4	Biogeográfia	3.8	Astrobiológia	3.6
	1 QUATERNARY		1 GEOLOGICA CARPATHICA	12	1 28TH EUROPEAN CONFERENCE ON SOLID-STATE TRANSDUCERS (EUROSENSORS 2014)	8	1 BIOGEOSCIENCES	3	1 GLOBAL ECOLOGY AND	1
	2 JOURNAL OF GEOSCIENCES	13	2 JOURNAL OF INTERNATIONAL	9	2 2012 IEEE SENSORS PROCEEDINGS	4			2 JOURNAL OF BIOGEOGRAPHY	1
	3 REVIEWS OF ARCHAEOLOGICAL SCIENCE	11	3 JOURNAL OF EARTH	6	3 CHANGE	3			3 LANDSCAPE ECOLOGY	1
	4 PALAEOGEOGRAPHY PALAEOCLIMATOLOGY	9	4 GEOMORPHOLOGY	5	4 GLOBAL SPATIAL SOIL INFORMATION SYSTEM	3			4 PROGRESS IN PHYSICAL GEOGRAPHY	1
	5 PALAEOGEOLOGY	5	5 QUATERNARY	3	5 19TH EUROPEAN SIGNAL PROCESSING CONFERENCE (EUSIPCO-2013)	2			5 REMOTE SENSING OF ENVIRONMENT	1
Tiszta kognitív hálózat	Negyedidőszak, vulkanológia.	22.1	Németi geológia	11.8	Eghajlátváltozás, paleogeográfia	10.5	Klimatológia, hidrologia	6.5	Távérzékelés	6.3
	1 QUATERNARY SCI REV	3074	1 EARTH PLANET SC LETT	317	1 GLOBAL CHANGE BIOL	900	1 INT J CLIMATOL	657	1 REMOTE SENS ENVIRON	197
	2 QUATERN INT	2560	2 CHEM GEOL	317	2 PALAEOGEOGR PALAEOCL	675	2 HYDROL	310	2 MON WEATHER REV	178
	3 TECTONOPHYSICS	1994	3 GEOCHIM COSMOCHEM AC	296	3 BIOGEOSCIENCES	469	3 CLIMATE	287	3 Q J ROY METEOR SOC	163
	4 J VOLCANOL GEOTH RES	1620	4 PALAEOGEOGR PALAEOCL	269	4 NATURE	460	4 B AM METEOROL SOC	174	4 IEEE T PATTERN ANAL	95
	5 PALAEOGEOGR PALAEOCL	1591	5 LITHOS	247	5 SCIENCE	440	5 THEOR APPL CLIMATOL	173	5 BIOL CONSERV	93

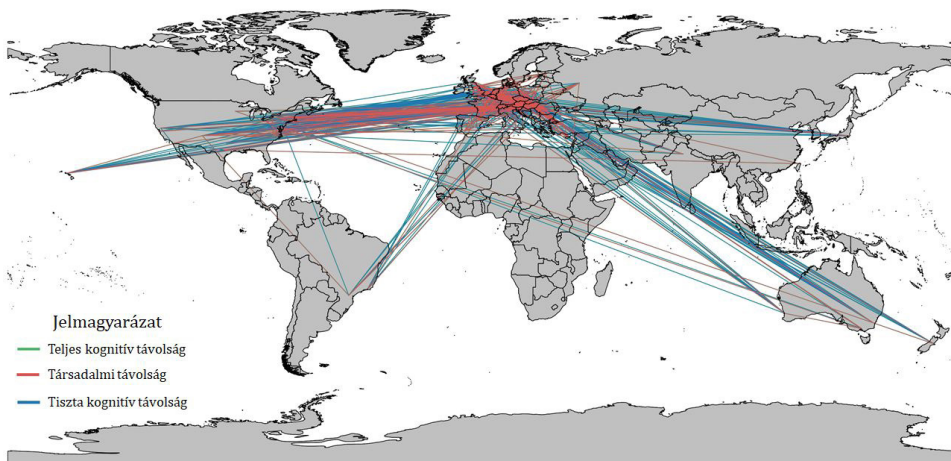
28. táblázat: Az öt legnagyobb szerzői modul öt leggyakoribb folyóirata és ez alapján a modulok tématerület-elnevezései mindhárom távolságtípus esetében a természetföldrajz területén

A természetföldrajz területén kirajzolódó főbb tématerületek a következők voltak: negyedidőszak-kutatás, klimatológia vagy éghajlatlan, távérzékelés, geológia (28. táblázat). A közgazdaságtan területén az egészség-gazdaságtanhoz kötődő folyóiratok voltak gyakoriak. Ennek egyik oka feltételezhetően a WoS Core Collection mint adatforrás használata volt. Az egészség-gazdaságtan mellett még olyan klasszikus tématerületek rajzolódtak ki, mint pénzügy, politikai gazdaságtan, agrárgazdaságtan, ökonometria és játékelmélet.

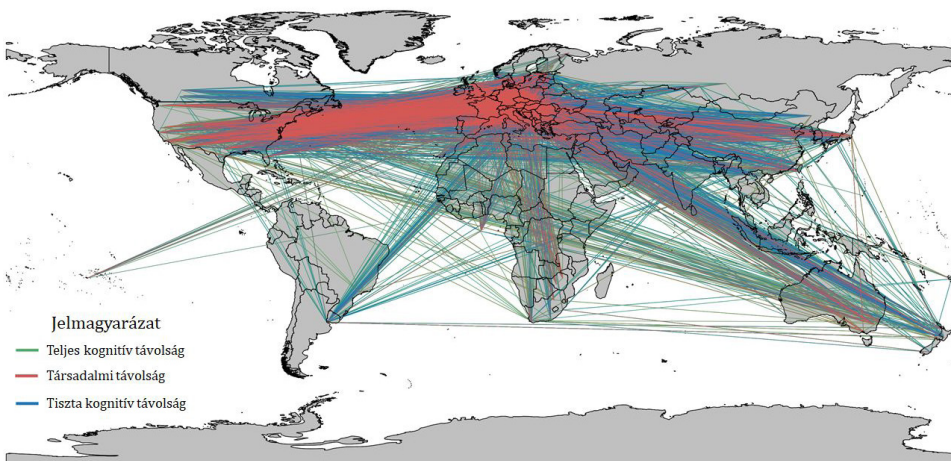
Bár a modularitás a hálózat szerkezete alapján különítette el a szerzői csoportokat, a hierarchikus klaszteranalízishez hasonló tématerületek rajzolódtak ki. Megjegyezhető, hogy a modularitás esetében némileg jobban elkülönültek egymástól a témacsoportok. A témacsoportok hasonlósága ebben az esetben is azzal magyarázható, hogy az együttműködések létrejöttéhez fontos az abban részt vevő kutatók hasonló tudásbázisa. Ezt az is alátámasztja, hogy a két csoportképző eljárás különböző elven alakította ki a szerzői csoportokat, azonban az azok alapján megjelenő tématerületek hasonlóak voltak.

7.3. A társadalmi, teljes és tiszta kognitív távolságtípusok a földrajzi térben

A 6.3. fejezetben bemutattam a társadalmi, teljes és tiszta kognitív hálózatok térképre vetítését oly módon, hogy a hálózatokat a szerzők települési affiliációi révén jelenítettem meg. A társadalmi, teljes és tiszta kognitív hálózatok települések közötti kifizítése révén lehetőségem volt megvizsgálni a földrajzi tér hatását a társadalmi és kognitív kapcsolatokra. Ezt három léptékben, világ-, Európa-szinten és a Magyarországgal szomszédos országok szintjén tettem meg.



28. ábra: A társadalmi, a teljes és a tiszta kognitív hálózatok a földrajzi térben a közgazdaságtan területén

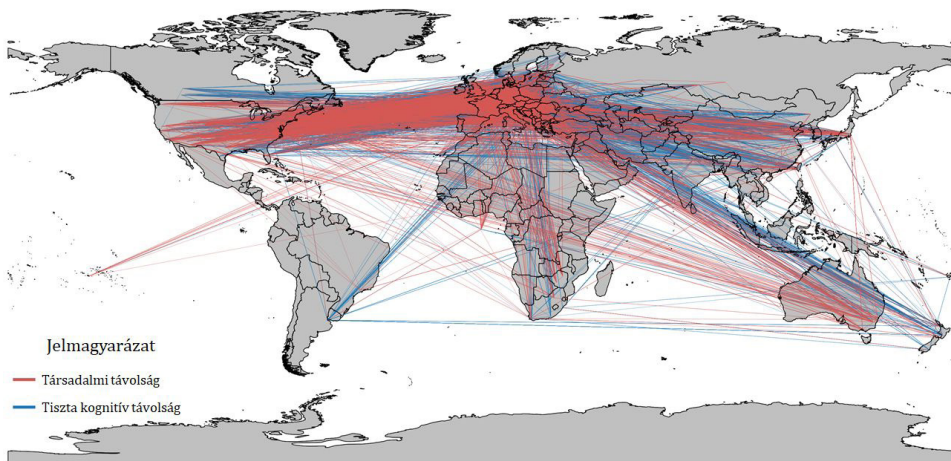


29. ábra: A társadalmi, a teljes és a tiszta kognitív hálózatok a földrajzi térben a természetföldrajz területén

A 28. és 29. ábra a három távolságtípus által meghatározott hálózatokat ábrázolja világszinten. A természetföldrajzi hálózatokban a jóval több él és csúcs sűrűbb hálózati képet mutat. A 7.2.1. alfejezet 19. táblázata szerint is látható volt, hogy míg a szerzők száma valamivel több a természetföldrajz területén, addig az élek számában már nagyságrendi különbség volt a természetföldrajz javára. A jóval sűrűbb természetföldrajzi hálózatok a földrajzi térben is szélesebb körű kiterjedést mutatnak, például az afrikai kontinens is bekapcsolódik a társadalmi és kognitív térbe. Doktori kutatásomban a földrajzi távolság vizsgálatára csupán a térképre vetítések révén került sor. A kutatásban használt minta települési szinten mutatja meg a magyar szerzők társszerzői

és kognitív kapcsolatait. Látható, hogy mind a két dimenzió, a társadalmi és a kognitív kapcsolatok is földrajzilag az egész világra kiterjedtek. Amikor két olyan település között található él, amelyek egyike sem Magyarországon található, annak oka a hálózatokra jellemző klaszterezettség. A társszerzői hálózatok esetében például, ha egy cikket három szerző publikált, akik közül csupán egy szerző volt magyar, akkor az a publikáció (egyebek mellett) megfelelt az alapkorpuszba kerülés feltételeinek. A két külföldi szerző között az összeköttetést a magyar szerző teremti meg a társszerzőség révén, ez pedig a hálózatban úgy jelenik meg, hogy a magyar szerző és a két külföldi szerző között lesz egy-egy él, de a két külföldi szerző között is lesz kapcsolat, azaz a hálózatban megjelenik egy háromszög. Háromnál több szerzős publikációkban még több háromszög rajzolódik ki.

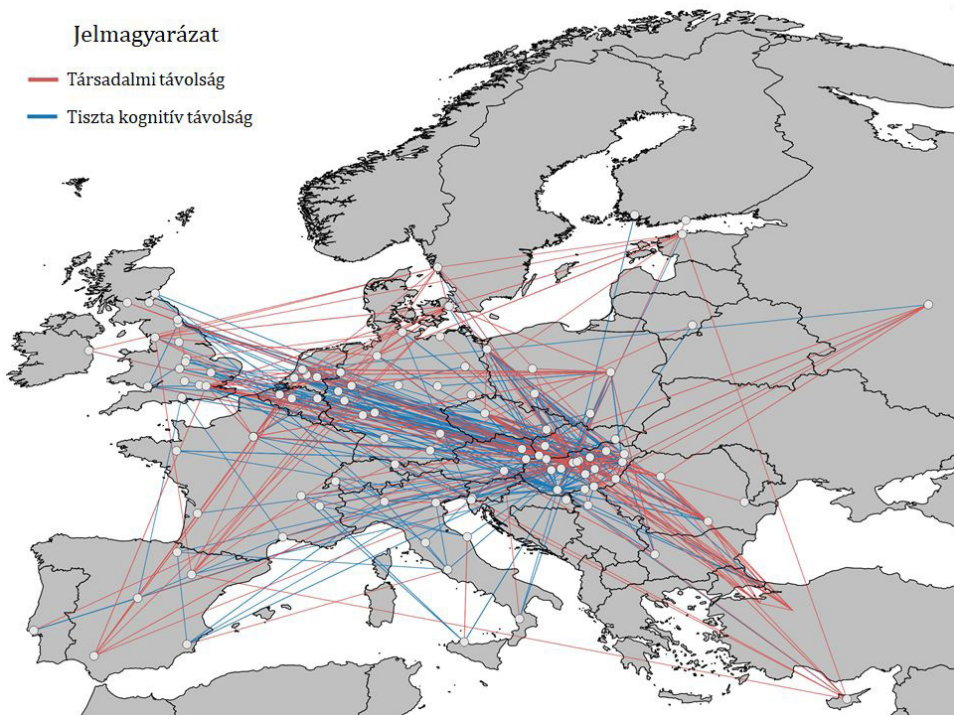
Az alapkorpusz meghatározása egy erős földrajzi paramétert tartalmazott (a publikációk szerzői között legalább egy magyar affiliációval rendelkezőnek kellett lennie), emiatt a hálózatok földrajzi vetülete Magyarország-központú volt. A természetföldrajz területén ábrázolt hálózatok olyan sok éllel rendelkeznek, hogy a térbeli összefüggések nem láthatók jól (29. ábra). A 30. ábra így a 29. ábra élei közül csak azokat tartalmazza, amelyek értékei az élsúlyok mediánja felett volt. A társadalmi kapcsolatok mediánja mindkét tudományterületen 1, azaz az adatsor maximuma volt. Ennek hátterében az áll, hogy a vizsgált öt éves időszak alatt a szerzők többségének csak egy publikációja volt a mintában, és ha ez társszerzőségben valósult meg, akkor az abban részt vevő szerzők között 1 volt a társadalmi hasonlóság. A 30. ábrán a teljes kognitív hálózat kapcsolatai sem kerültek ábrázolásra, mivel ezek a kapcsolatok megjelennek a térképen vagy a társadalmi, vagy a tiszta kognitív kapcsolatok településközi összeköttetései révén. Így egy valamivel letisztultabb képet kaptam. A 30. ábra Magyarország jelentősebb (a kapcsolatok mediánszűrése miatt) társadalmi és tiszta kognitív kapcsolatait mutatja a természetföldrajz területén. Mivel Európában olyan sok a kapcsolat, hogy az nem látszódik, így ez a térkép Magyarország európai kontinensen kívüli kapcsolatait mutatja jobban. A 28., 29. és 30. ábrák alapján hazánk társadalmi és kognitív kapcsolatai behálózzák a világot, és kapcsolatban állunk a kibocsátás és idézettség szempontjából nemzetközi szinten jelentős országokkal (pl. Egyesült Államok, Kína, Ausztrália, Japán). Kijelenthetjük, hogy mind a két vizsgált tudományterületen hazánk bekapcsolódik a globális tudományos együttműködési hálózatba, valamint a nemzetközi tudásbázist is felhasználja kutatásai során.



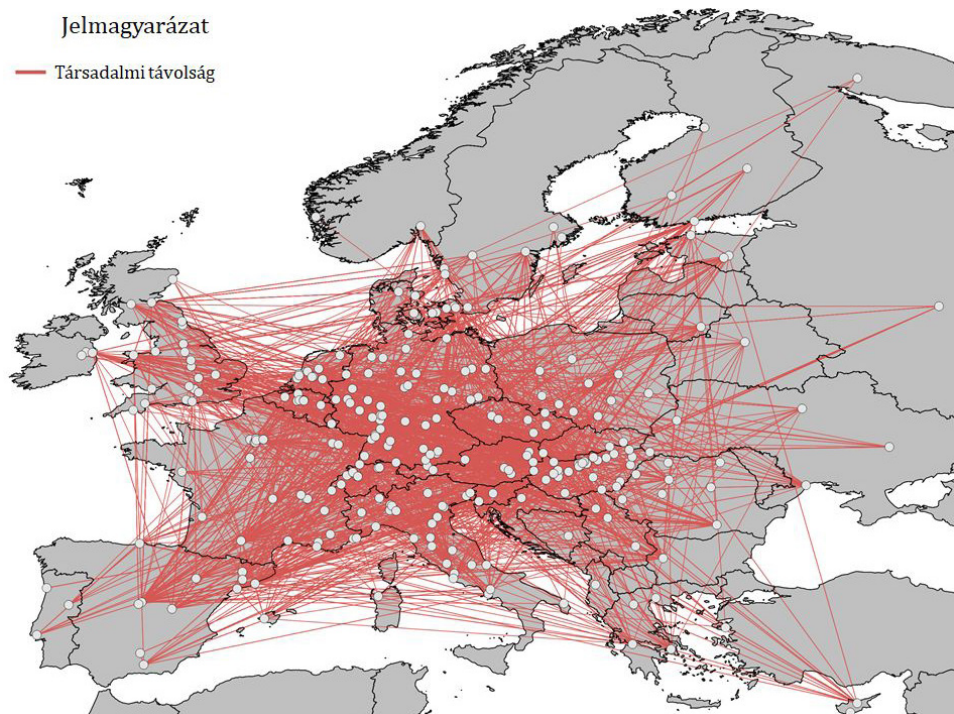
30. ábra: A társadalmi és a tiszta kognitív hálózatok medián feletti élsúllyal rendelkező kapcsolatai a földrajzi térben a természetföldrajz területén

Ezt követően csak az Európán belül megvalósuló kapcsolatok kerültek ábrázolásra az egyes hálózatokban. A teljes kognitív kapcsolatokat a jobb átláthatóság végett itt sem ábrázoltam. A 31. ábra a közgazdaságtan területén mutatja a társadalmi és a tiszta kognitív hálózatok medián feletti élsúllyal rendelkező kapcsolatait. A jelentősebb társadalmi kapcsolatok egész Európára jellemzők. Főbb együttműködő partnerországaink Németország, Egyesült Királyság, Belgium, Hollandia és a szomszédos országok, mint például Románia.

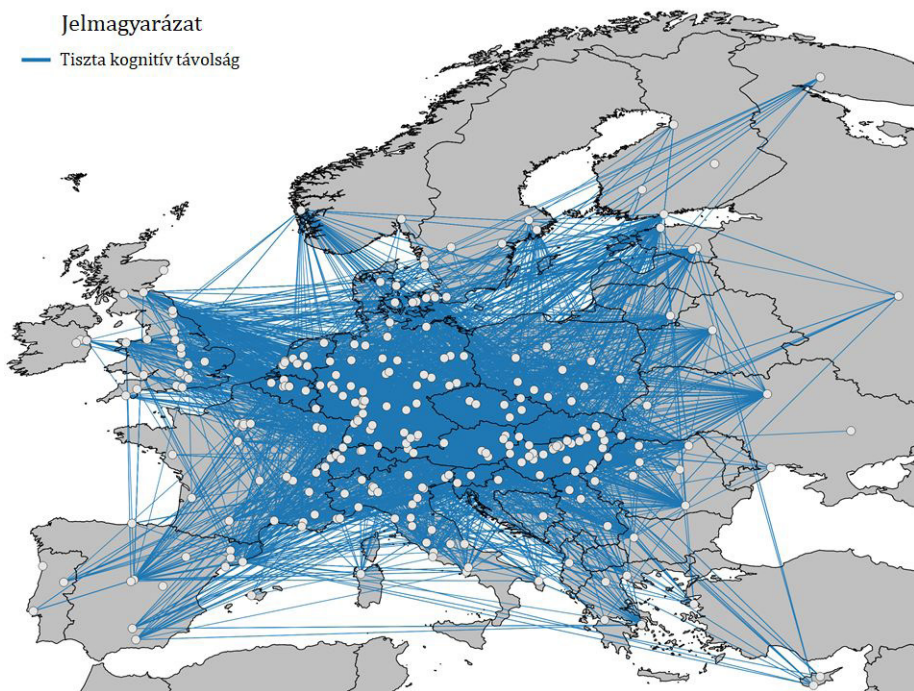
A társadalmi kapcsolatoktól mentes tudásbázison alapuló tiszta kognitív kapcsolatokban erőteljes nyugati irányultság látható. Ez azt jelenti, hogy a magyarországi szerzők a nyugat-európai kollégákkal hasonló irodalmi tételeket használtak fel a kutatásuk során. A közgazdaságtan területén az alkalmazott tudásanyag szempontjából Magyarország része a nyugat-európai mainstreamnek. Mindezek ellenére érdemes megjegyezni, hogy mivel az alapkörpusz a WoS-ból származott, az abban szereplő publikációk eleve részei a tudományos fővonulatnak. Ugyanakkor, ahogy azt az 5. fejezetben kifejtettem és sokan mások is megjegyzik, tudományometriai elemzésekre ez az egyik leggyakrabban használt és leginkább használható citációs adatbázis (Bar-Ilan, 2008; Rafols, 2017 STI plenáris előadás).



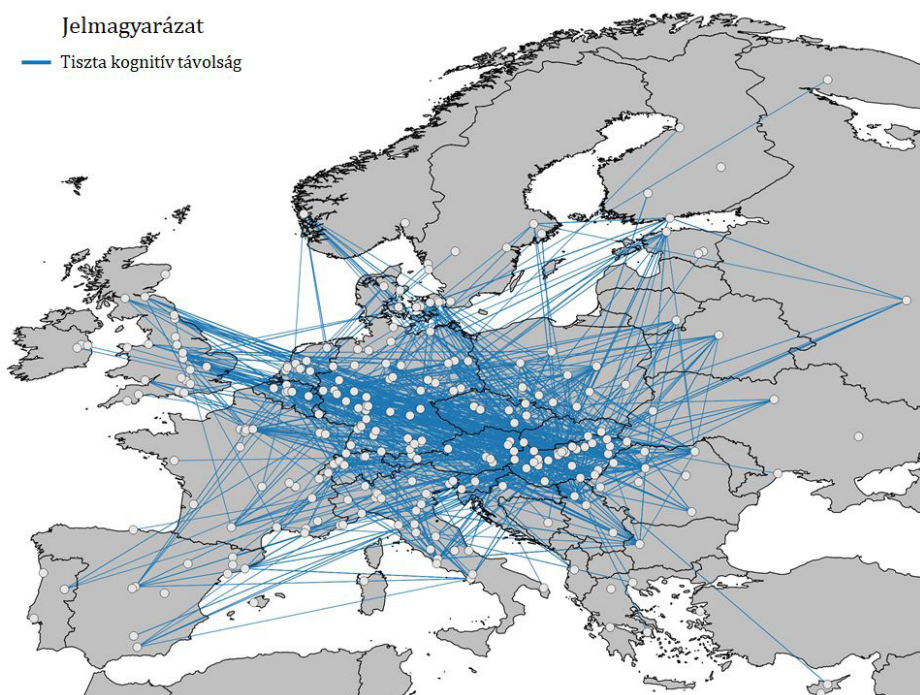
31. ábra: A társadalmi és a tiszta kognitív hálózatok medián feletti élsúlyal rendelkező kapcsolatai Európában a közgazdaságtan területén



32. ábra: A társadalmi hálózat medián feletti élsúlyal rendelkező kapcsolatai Európában a természetföldrajz területén



33. ábra: A tiszta kognitív hálózat medián feletti élsúllyal rendelkező kapcsolatai Európában a természetföldrajz területén



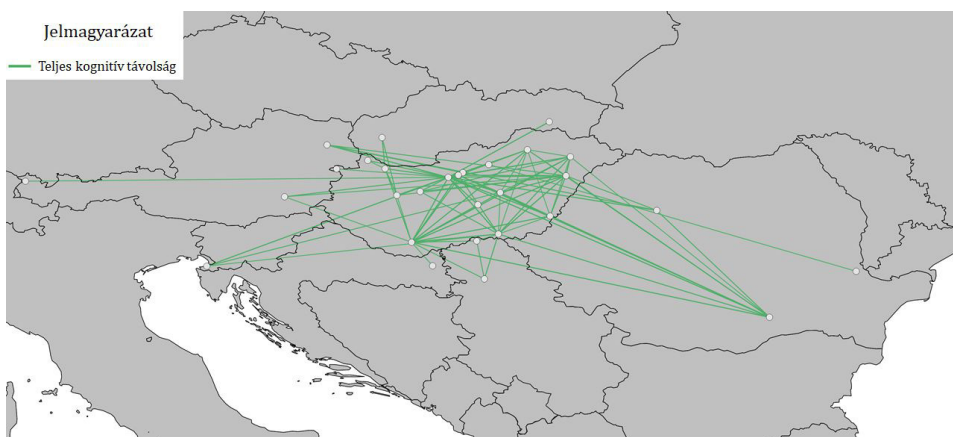
34. ábra: A tiszta kognitív hálózat kapcsolatainak legerősebb 10%-a Európában a természetföldrajz területén

A 32. és 33. *ábra* a természetföldrajz területén mutatja be a társadalmi és a tiszta kognitív hálózatok jelentős, medián feletti élsúllyal rendelkező kapcsolatait. A természetföldrajz területén a társadalmi és a kognitív dimenzióban is számos kapcsolat található Európában, ezért ezeket külön-külön ábrázoltam. A kognitív kapcsolatoknak az előzőekben a közgazdaságtan területén megfigyelt nyugati irányultsága a természetföldrajz területén nem jelentkezik. Magyarország társadalmi és kognitív kapcsolatai egész Európán belül megjelennek. Ha azonban a kognitív kapcsolatok legerősebb 10%-át nézzük (34. *ábra*), akkor a természetföldrajz területén is megjelenik a nyugati irányultság, azaz a magyar és a nyugat-európai szerzők hasonló publikációkra hivatkoztak.

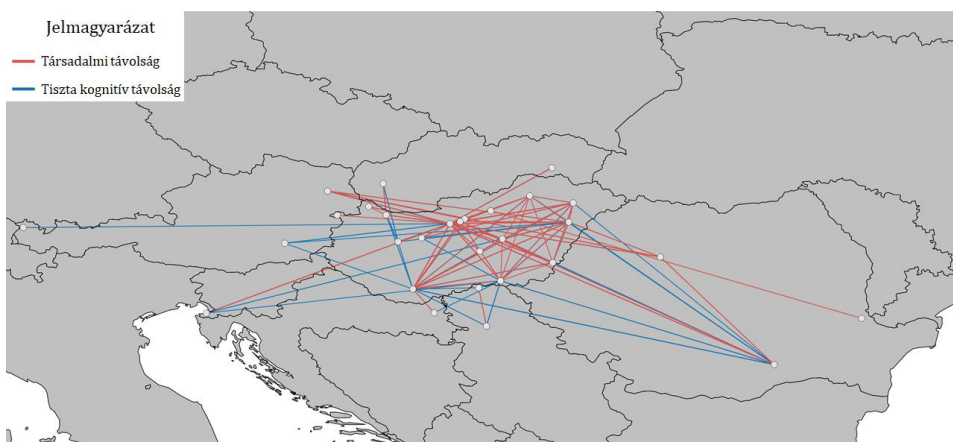
A 35., 36. és 37. *ábra* Magyarország és a hazánkkal szomszédos országok között mutatja a kognitív és a társadalmi kapcsolatokat a két tudományterületen. Az ábrákon csak azok az élek kerültek megjelenítésre, melyekben mindkét szerző Magyarországon vagy Magyarországgal szomszédos országban található települést jelölt meg a publikáción affiliációként. A közgazdaságtan területén minden kapcsolat megjelenítésre került, míg a természetföldrajz esetében csak a medián feletti élsúllyal rendelkező kapcsolatok.

A 35. *ábrán* kizárólag a teljes kognitív kapcsolatokat, míg a 36. *ábrán* a társadalmi és a tiszta kognitív kapcsolatokat ábrázoltam. Láthatjuk, hogy minden településpár esetében, ahol teljes kognitív kapcsolat volt, az tiszta kognitív és/vagy társadalmi kapcsolat meglétét is jelentette. Egy településpár között azért lehet társadalmi és tiszta kognitív kapcsolat is, mivel a szerzői hálózatokat vetítettem a térképekre. Egy településpár esetében két szerző között lehet társadalmi kapcsolat, míg másik két szerző között, akik ugyanazon településpárt jelölték meg affiliációnak, lehet tiszta kognitív kapcsolat.

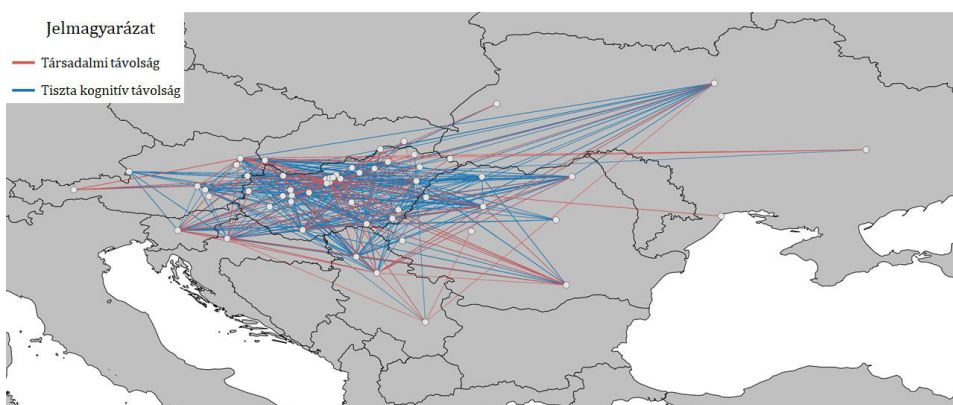
Magyarország és az azzal szomszédos országok földrajzi léptékében is a természetföldrajz mutatott sokkal több kapcsolattal rendelkező hálózatokat. A közgazdaságtan területén (36. *ábra*) az országon belüli kapcsolatok közül a társadalmi kapcsolatok jellemzőbbek voltak, mint a tiszta kognitív kapcsolatok. A természetföldrajz területén ilyen mintázat nem volt látható, ott mind a társadalmi, mind a tiszta kognitív kapcsolatok az országon belül és a szomszédos országokkal is sűrűek voltak.



35. ábra: Magyarország és szomszédai közötti teljes kognitív kapcsolatok a közgazdaságtan területén



36. ábra: Magyarország és szomszédai közötti társadalmi és tiszta kognitív kapcsolatok a közgazdaságtan területén



37. ábra: Magyarország és szomszédai közötti medián feletti élsúllyal rendelkező társadalmi és tiszta kognitív kapcsolatok a természetföldrajz területén

A térképek segítségével szemléltettem, hogy Magyarország társadalmi és tiszta kognitív, illetve társszerzői kapcsolatai mind a közgazdaságtan, mind a természetföldrajz területén az egész világra kiterjednek. A hazai kutatók jellemzően a magyar kutatókkal, a szomszédos országok kutatóival és világszinten a kibocsátás és idézettség szempontjából jelentős országok kutatóival (pl. Egyesült Államok, Kína, Ausztrália, Japán) működnek együtt. A tiszta kognitív kapcsolatok is behálózzák az egész világot, azaz a magyar kutatók és a nemzetközi kutatók által használt irodalmak hasonlóak, a magyarországi kutatók hozzáférnek a tudományos mainstream szakirodalomhoz, ismerik és használják azt.

7.4. Az empirikus eredmények összefoglalása

Az értekezés empirikus részében egyrészt hálózathálózati és hálózatelemzési módszerekkel hasonlítottam össze a három hasonlósági mátrixot, másrészt a három mátrixból létrehozott hálózatot kifestettem a földrajzi térbe és megvizsgáltam a belső és külső tér átfedéseit. A disszertációban meghatározott tiszta kognitív távolság a társadalmi kapcsolatoktól mentes tudásbázist hivatott megragadni.

A hasonlósági mátrixok értékeit (7.1.1. alfejezet), illetve a hálózatok esetében az élsúlyokat figyelembe vevő QAP korreláció (7.2.2. alfejezet) módszer bizonyította a társadalmi és tiszta kognitív távolságtípusok elkülönítését. E két módszer alapján megállapítható volt, hogy a társadalmi komponens meghatározóbb faktor a teljes kognitív távolságban.

A hálózatok szerkezeti sajátosságain alapuló összehasonlítások (7.2.1 alfejezet) a két kognitív távolságtípus – a teljes és tiszta kognitív távolság – mutatott nagyfokú átfedést. Ezek a módszerek az összehasonlítások vagy nem, vagy csak részben vették figyelembe az élsúlyokat, így ezek csak a szerzők közötti kapcsolatokat vizsgálták. Jelen kutatásban azonban az egyes távolságtípusok elkülönítésének bizonyításakor épp a hasonlósági mátrixok értékei, illetve a hálózatokban az élsúlyok voltak a lényegesek.

A három távolságmátrix (társadalmi, teljes kognitív és tiszta kognitív) szerzőit hierarchikus klaszteranalízis (7.1.2. alfejezet) és hálózati modularitás (7.2.3. alfejezet) segítségével csoportosítottam. Céлом az volt, hogy igazoljam a tiszta kognitív távolság elkülönítését a létrejött szerzői csoportok különbözősége alapján is. Ehhez a módosított Rand-indexet alkalmaztam, mely a létrejött csoportok szerzői közötti átfedés mértékét mutatta meg. A hierarchikus klaszteranalízis során a tiszta kognitív távolságmátrix szerzői csoportjai teljesen elkülönültek a másik két távolságmátrix csoportjaitól, míg a társadalmi és a teljes kognitív távolságmátrixok csoportjai között nagy volt az átfedés.

Ez igazolta egyrészt a tiszta kognitív távolság társadalmi hatástól való elkülönítését, a magas társadalmi és teljes kognitív hasonlóság pedig az erős társadalmi komponensre volt visszavezethető. A hálózati modularitás alapján meghatározott csoportok esetében a két kognitív hálózat között tapasztaltam nagy átfedést. A társadalmi és a tiszta kognitív komponens itt is elkülönült egymástól. A két módszerből adódó eredmények ismét alátámasztották azt, hogy amikor a mátrixok cellaértékei alapján történt a csoportok lehatárolása, akkor a teljes kognitív és a társadalmi távolságmátrix hasonlított inkább egymásra. A hálózati modularitás alkalmazásakor a hálózat szerkezete, tehát a kapcsolatok léte erősebb tényezőnek bizonyult a kapcsolatok erősségénél, ugyanis a módszer algoritmus, bár figyelembe veszi az élsúlyokat a hálózati csoportok kialakításakor, mégis a hálózat élein halad végig, azokból indul ki.

A legnagyobb öt szerzői csoport tématerületeit is megvizsgáltam. Megnéztem, hogy a kognitív távolság és a társadalmi távolság alapján meghatározott tématerületek mennyire hasonlítanak egymásra. Mind a tudományterületek között, mind pedig a két csoportképző módszerrel arra az eredményre jutottam, hogy hasonló tématerületek jelennek meg az egyes távolságdimenziók mentén. Ez azt jelenti, hogy a társadalmi és a kognitív távolság esetében is a szerzők hasonló tudásbázis alapján kerülnek egy csoportba. Ez alátámasztja azt, hogy az együttműködések szerveződésekor a hasonló érdeklődési terület elősegíti az együttműködések létrejöttét, az együttműködések pedig elősegítik a résztvevők tudásbázisának közeledését.

A három hálózatot a földrajzi térbe vetítettem a szerzők affiliációi révén (7.3. fejezet). A csúcsok a települések lettek, míg az élek maradtak a hasonlósági mátrix kapcsolati értékei. Látható volt, hogy a két kognitív hálózat sokkal sűrűbb hálózatot mutatott, mint a társadalmi hálózat, viszont a társadalmi hálózat kapcsolatai erősebbek voltak, már az adatsor mediánja megegyezett az adatsor maximumával. Három földrajzi léptékben – világ, Európa, Magyarország és az azzal szomszédos országok – vizsgáltam hazánk társszerzői és tiszta kognitív kapcsolatait. A természetföldrajz hálózatai mindhárom esetben sokkal sűrűbbek voltak. Mind a két tudományterületen hazánk társadalmi kapcsolatai egyrészt a földrajzilag közel eső, hazai és a környező országok szerzőivel valósultak meg, másrészt Európa- és világszinten jelentős publikációs kibocsátó és magas idézettséggel rendelkező országok kutatóival. A tiszta kognitív kapcsolatok is követték ezeket a trendeket. Magyarország hasonló tudásanyagra épít, mint a világ kutatói. A közgazdaságtan területén hazánk európai kapcsolatait vizsgálva azt találtam, hogy a tiszta kognitív kapcsolatok erős nyugati irányultságot mutattak, ez a természetföldrajz területén csak a legerősebb 10%-ba tartozó tiszta kognitív kapcsolatok esetében volt így. Ez a nyugati irányultság a közgazdaságtan világszintű kapcsolataiban is jelentkezett.

A két kiválasztott tudományterület hasonló képet mutatott a különböző távolságtípusok összehasonlításakor. Megállapítható, hogy a természetföldrajz esetében a társadalmi komponens dominánsabb, de a különbség a két terület között nem volt jelentős. A természetföldrajz területén tapasztalt társadalmi komponens meghatározóbb jelenléte visszavezethető arra a tényre, hogy a természettudományok területén gyakoribbak a társszerzőségek. A jelen vizsgálatba vont publikációs halmaz esetén a társszerzős cikkek aránya a természetföldrajzi és a közgazdaságtani cikkekre vonatkozóan 89% és 61%, míg az egy cikkre jutó átlagos szerzőszám 6, illetve 3 volt.

8. Összefoglalás

Napjainkban a tudás létrehozása egyre gyakrabban kutatócsoportok együttes munkájához kötődik, ami a társszerzős publikációk arányának növekvő tendenciájában is megmutatkozik. Ez a növekvő arány nem csupán a hagyományosan is magasabb átlagos szerzőszámokkal rendelkező természettudományi területeken, hanem a társadalomtudományokban is megfigyelhető.

Doktori kutatásomban az egyre nagyobb jelentőséggel bíró tudományos együttműködések vizsgáltam, különös tekintettel a létrejöttüket meghatározó, szereplők között fennálló távolságdimenziók szempontjából. A tudományos együttműködések a társszerzőségeken keresztül azonosítottam, szem előtt tartva, hogy a két fogalom nem ugyanazt takarja: ezt a kérdéskört a 3. fejezetben mutattam be.

A kutatók munkásságuk során önálló publikációik mellett számos társszerzőségben vesznek részt. Ezek a társszerzőségek egy hálózatot alkotnak, melynek vizsgálata hálózatelemzési módszerekkel lehetséges. A társszerzői hálózatok tulajdonképpen kapcsolathálózatok, mivel személyek közötti tényleges relációk alapján jönnek létre. Az élet számtalan területe megragadható hálózatokkal. A hálózattudomány térhódítása feltárta, hogy a különböző hálózatok (pl. internet, baráti kapcsolathálózatok, fehérjehálózatok) néhány alapvető tulajdonságukban hasonlítanak egymásra (pl. a hálózat méretéhez képest rövid átlagos úthossz, csomópontok megjelenése), ez a társszerzői hálózatokra is igaz.

Az értekezésben bemutattam, hogy az egyes, távolsághoz kapcsolódó tércategóriák hálózatokon is értelmezhetők. Mivel a hálózatok esetében a fókusz a kapcsolatok létén és erősségén van, azok csupán a valóság egyszerűsített leképezései, a bemutatott tércategóriák értelmezése (4.1. fejezet) a regionális tudományi tartalomhoz képest módosul. Az egyes tércategóriák szerepe a hálózatokon történő áramlás esetén válik szembetűnővé, például a *szomszédság* esetében egy vélemény vagy járvány terjedésekor meghatározó szerepe van a szomszédok állapotának. A hálózatoknál az egy lépés távolságra található csúcok tekinthetők egy csúcs elsőfokú szomszédainak, míg a külső térben a területi objektumok akkor szomszédosak egymással, ha közöttük a távolság 0, és nem egyeznek egymással. A *határ* is az egyes modulok között a hálózaton történő áramláskor válik jól megfoghatóvá: egy csoporton belül, ahol sok a csoporttagok között a kapcsolat, az információ gyorsan elterjed, viszont mivel a csoporton kívülre kevés kapcsolat mutat, a határ elválasztó és összekötő szerepe egyszerre mutatkozik meg.

A kutatói együttműködések létrejöttében számos tényező játszik szerepet. A szereplők között egy-egy tényező mentén meghatározható a távolság, mely minél kisebb, annál jobban hasonlítanak egymáshoz az együttműködésben potenciálisan résztvevők. Emiatt a kisebb távolságértékek kedveznek az együttműködések létrejöttének, de ez csak egy bizonyos pontig igaz: a túlzott hasonlóság gátolhatja is az együttműködések létrejöttét. A szakirodalom rendszerezése során a különböző tényezőket három távolságtípus-csoportba (*földrajzi, társadalmi és kognitív távolság*) soroltam.

A társszerzőségek létrejöttekor a társadalmi (nyelvi, történelmi, kulturális, gazdasági) tényezők befolyásolják az egyének döntéseit, hogy kivel működnek együtt. A kognitív távolságot a hivatkozások hasonlóságán keresztül tártam fel, azon feltételezés alapján, hogy a hasonló hivatkozások hasonló kutatási témát rejtenek. Ugyanakkor be kell látni azt, hogy a fentebb felsorolt társadalmi tényezők azt is befolyásolják, kire történik a hivatkozás. A földrajzi közelség sok esetben ráerősít erre: a földrajz első törvényében is kifejeződik a földrajzi közelség során megjelenő hasonlóság. Látható tehát, hogy az elkülönített és vizsgált távolságtípusok, bár különböző dimenziók mentén írják le a távolságot, nem függetlenek egymástól. Doktori kutatásomban én mégis arra tettem kísérletet, hogy ezeket egymástól elkülönítsem és ezáltal ezek külön-külön elemezhetők legyenek.

Az együttműködések azok szerveződési szintjén, a szerzői szinten vizsgáltam. A hivatkozások hasonlóságát a szerzői bibliográfiai csatolás módszerén keresztül mértem, mely eljárás alkalmazásához szükség volt a hivatkozások publikációk szintjéről szerzői szintre történő vetítésére. Míg két publikáció között akkor áll fenn bibliográfiai csatolás, ha ugyanazokra a publikációkra hivatkoztak, addig a szerzői szinten ugyanez a szerzők között kerül azonosításra. A publikációk szintjéről a szerzői szintre vetítésnél az adott publikáció minden egyes szerzőjéhez hozzárendeltem a publikáció összes hivatkozását. Így, amikor egy társszerzős cikk hivatkozásait a szerzői szintre vetítettem, akkor a szerzők között teljesen megegyeztek a hivatkozások. Ez ebben az esetben azt jelentette, hogy a publikáció társszerzői között a bibliográfiai csatoláson keresztül (amennyiben a szerzők ezzel az egy publikációval szerepeltek a mintában) teljes hasonlóság volt. Ez teljes kognitív egyezést jelentene, holott ennek oka csupán a társszerzőségben, azaz a társadalmi távolságkomponensben gyökerezik. Ebből a példából látható, hogy a társszerzőség miatt lehet kognitív hasonlóság a szerzők között. Ugyanakkor kognitív hasonlóság a társszerzőség nélküli közös hivatkozáshasználatból is adódhat. Az előbbi esetet *társadalmi komponensnek*, az utóbbit pedig *tiszta kognitív komponensnek* neveztem el. A szerzők között alkalmazott szerzői bibliográfiai csatolás eredménye mind a két komponenst tartalmazza, ezért ennek a *teljes kognitív távolság* elnevezést adtam. A társadalmi komponens azonosítására a társ-

szerzőségek alkalmasak, mivel ezeken keresztül jelenik meg. A tiszta kognitív komponenst a teljes kognitív távolságot leíró hasonlósági mátrix és a társadalmi komponenst megragadó társszerzőségi hasonlósági mátrix különbségének abszolút értékeként határoztam meg. A hasonlósági mátrixokat a Salton-index segítségével hoztam létre, így azok cellaértékei 0 és 1 közötti intervallumba estek. Ezáltal gyakorlatilag normalizáltam a nyers mátrixokat, és így már ki lehetett egymásból vonni azokat. Az adatok abszolút értékre történő transzformációjára a negatív hasonlósági értékek kiszűrése miatt volt szükség, mivel számomra a kapcsolat erőssége és nem az iránya volt fontos a tiszta kognitív távolság meghatározásakor.

A konceptuális modellben elkülönített komponensek igazolására az értekezés második felében két tudományterületen – közgazdaságtan és természetföldrajz – 2010 és 2014 között megjelent, WoS-ben elérhető publikációk közül azokat vizsgáltam, melyeknek legalább egy magyarországi affiliációval rendelkező szerzőjük volt. A számításokhoz az R, a vizualizációhoz a Gephi szoftvert használtam.

A létrehozott hasonlósági mátrixokat hálózatifüggetlen, a hasonlósági mátrixokból képzett hálózatokat pedig hálózatelemzési módszerekkel vizsgáltam. Első közelítésben összehasonlítottam a teljes kognitív, tiszta kognitív és társadalmi hasonlósági mátrixokat az alapján, hogy az egyes hasonlóságerősségi szintekbe a cellák hány százaléka tartozik. A társadalmi és a tiszta kognitív komponens elkülönülése jól megfigyelhető volt. A hálózatelemzési módszerek közül a QAP korreláció szintén alátámasztotta a távolságtípusok elkülönítését és a társadalmi komponens meghatározóbb jelenlétét mutatta a teljes kognitív távolságon belül. A többi hálózatelemzési módszer elsősorban a hálózat szerkezete alapján hasonlította össze az egyes hálózatokat. Ekkor a teljes és a tiszta kognitív hálózatok hasonlósága volt nagyobb fokú. A két kognitív hálózat hasonlósága a hasonló kapcsolatokon (hivatkozások) való alapulásuk miatt volt megfigyelhető, ezek a hálózatok a társszerzői hálózathoz képest jóval több kapcsolattal rendelkeztek. Mivel az alapvető hálózati paraméterek összehasonlításakor a kapcsolatok léte sokkal nagyobb szerepet kapott a kapcsolatok erősségénél, ezeknél a módszereknél a két kognitív hálózat hasonlított egymásra jobban. A vizsgálatomban azonban a kapcsolatok erőssége volt a lényeges, ez tükrözte a távolságot az egyes szereplők között.

Hierarchikus klaszteranalízis és hálózati modularitás segítségével csoportosítottam a vizsgálatban részt vevő szerzőket a társadalmi, teljes és tiszta kognitív távolságmátrixokban és hálózatokban. A létrejött csoportok szerzői közötti átfedéseket a módosított Rand-index segítségével hasonlítottam össze. A társadalmi és a tiszta kognitív komponensek elkülönülése a szerzői csoportok szintjén is megmutatkozott mind a két módszer esetében. A hierarchi-

kus klaszteranalízis a csoportok kialakításakor a távolságmátrix cellaértékeit vette figyelembe, ekkor a teljes kognitív távolság a társadalmi távolságkomponenssel mutatott nagyobb hasonlóságot. Ezzel szemben a hálózati modularitás eredményeinek összehasonlításakor a két kognitív hálózat szerzői csoportjai hasonlítottak egymásra jobban, hiszen ekkor a hálózat szerkezete volt az, ami leginkább megszabta a modulok kialakulását. Az alkalmazott fast greedy algoritmus ugyan figyelembe veszi az élsúlyokat a modulok kialakításakor, azonban a hálózat élein halad végig, emiatt a kapcsolatok léte erősebb tényező az élsúlyoknál.

A létrejött legnagyobb szerzői csoportok tématerületeinek hasonlóságát is megnéztem. Mindkét módszer és tudományterület esetében hasonló tématerületeket találtam a társadalmi és a kognitív távolságoknál, azaz a csoportok szerveződését meghatározta a szerzők tudásbázisa. Míg az együttműködések kialakulásakor a hasonló tudásbázis előny, addig az együttműködésekben történő részvétel elősegíti a résztvevők tudásbázisának közeledését.

A földrajzi dimenziót a társadalmi és kognitív dimenziók földrajzi térben történő vetítésével vizsgáltam. A meghatározott teljes és tiszta kognitív, valamint társadalmi hálózatokat kifeszítettem a települési szintre, így láthatóvá váltak a társadalmi és kognitív dimenziók földrajzi összefüggései. Mind a két vizsgált tudományterületen hazánk társadalmi és kognitív kapcsolatai egyrészt a földrajzilag közel eső területekkel, másrészt a nemzetközileg magas publikációs kibocsátó és magas idézettségmutatókkal rendelkező országokkal (Németország, Egyesült Államok, Kína, Ausztrália, Japán, Hollandia és Belgium) valósultak meg. A tiszta kognitív kapcsolatok esetében megfigyelhető volt egy jellemzően nyugati irányultság a társadalmi kapcsolatokhoz képest: ez a közgazdaságtan területén jóval szembetűnőbbnek bizonyult, de a leg-erősebb tiszta kognitív kapcsolatok esetében a természetföldrajz területén is kirajzolódott.

A vizsgálatban szereplő természetföldrajzi és közgazdaságtani hálózatok, mátrixok a távolságtípusok különböző módon történő összehasonlításakor jellemzően megegyező képet mutattak. Ugyanakkor a természetföldrajz területén meglévő magasabb átlagos szerzőszám és a társszerzős publikációk magasabb aránya az eredményekben is tükröződött a dominánsabb társadalmi komponensen keresztül.

Doktori kutatásomban a tudományos együttműködési hálózatokban részt vevő kutatók kapcsolatait alapvetően két fő dimenzió (társadalmi és kognitív) mentén vizsgáltam. Az egyes dimenziókon elsődlegesen a távolság határozza meg, hogy a kutatók munkájuk során együttműködnek-e vagy sem. Emellett fontos, hogy az egyes dimenziók mentén egyáltalán létezik-e kapcsolat: ezt

a hálózati felfogás, a hálózatelemzési módszerek nyújtotta eredmények mutatták meg. Az egymással kapcsolatban álló szereplők közötti távolság, mely vizsgálatomban nagyobb szerepet kapott, a hálózatok élsúlyában, illetve a hasonlósági mátrixok cellaértékeiben fejeződött ki. Emiatt a végső következtetésnél azokat a módszereket vettem figyelembe, melyek elsődlegesen a kapcsolati értékek erősségét írták le. Ebből egyértelműen kirajzolódott a tiszta kognitív komponens teljes kognitív távolságtól történő elkülönítése.

Kutatásom értekezés egyik újdonsága, hogy a teljes kognitív távolságban azonosítottam és elkülönítettem a két komponenst egymástól. A létrejött tiszta kognitív komponens a társszerzőségek nélkül fennálló kognitív hasonlóság mérésére szolgál a szerzők között. Segítségével potenciális jövőbeli együttműködésben részt vevő kutatók kerülhetnek azonosításra.

Felhasznált irodalom

- Acosta, M., Coronado, D., Ferrándiz, E., & León, M. D. (2011). Factors affecting inter-regional academic scientific collaboration within Europe: The role of economic distance. *Scientometrics*, 87(1), 63-74. <https://doi.org/10.1007/s11192-010-0305-6>
- Aghaei Chadegani, A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ale Ebrahim, N. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *Asian Social Science*, 9(5), 18-26. <https://doi.org/10.5539/ass.v9n5p18>
- Ahlgren, P., & Jarneving, B. (2008). Bibliographic coupling, common abstract stems and clustering: A comparison of two document-document similarity approaches in the context of science mapping. *Scientometrics*, 76(2), 273-290. <https://doi.org/10.1007/s11192-007-1935-1>
- Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47. <https://doi.org/10.1103/RevModPhys.74.47>
- Angelusz, R., & Tardos, R. (2009). A kapcsolathálózati szemlélet a társadalom- és politikatudományban. *Politikatudományi Szemle*, XVIII/2., 29–57.
- Autant-Bernard, C., Billand, P., Frachisse, D., & Massard, N. (2007). Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies. *Papers in Regional Science*, 86(3), 495-519. <https://doi.org/10.1111/j.1435-5957.2007.00132.x>
- Autio, E., Hameri, A. P., & Nordberg, M. (1996). A framework of motivations for industry-big science collaboration: a case study. *Journal of Engineering and Technology Management*, 13(3-4), 301-314. [https://doi.org/10.1016/S0923-4748\(96\)01011-9](https://doi.org/10.1016/S0923-4748(96)01011-9)
- Babbie, E. (2004). *A társadalomtudományi kutatás gyakorlata*. Budapest: Balassi Kiadó.
- Barabási, A. L. (2003). *Behálózva. A hálózatok új tudománya*. Budapest: Magyar Könyvklub.
- Barabási, A.L., (2016). *A hálózatok tudománya*. Budapest: Libri.
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512. <https://doi.org/10.1126/science.286.5439.509>
- Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3-4), 590-614. [https://doi.org/10.1016/S0378-4371\(02\)00736-7](https://doi.org/10.1016/S0378-4371(02)00736-7)
- Barabasi, A. L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2), 101. <https://doi.org/10.1038/nrg1272>
- Bar-Ilan, J. (2008). Which h-index? – A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271. <https://doi.org/10.1007/s11192-008-0216-y>
- Batagelj, V. (1997). Notes on blockmodeling. *Social Networks*, 19(2), 143-155. [https://doi.org/10.1016/S0378-8733\(96\)00297-3](https://doi.org/10.1016/S0378-8733(96)00297-3)

- Beauchamp, M. A. (1965). An improved index of centrality. *Behavioral science*, 10(2), 161-163. <https://doi.org/10.1002/bs.3830100205>
- Beaver, D. D. (2001). Reflections on scientific collaboration (and its study): past, present, and future. *Scientometrics*, 52(3), 365-377. <https://doi.org/10.1023/A:1014254214337>
- Bell, G. G. (2005). Clusters, networks, and firm innovativeness. *Strategic Management Journal*, 26(3), 287-295. <https://doi.org/10.1002/smj.448>
- Benedek, G., Lublóy, Á., & Szenes, M. (2007). A hálózatelmélet banki alkalmazása. *Közgazdasági Szemle*, LIV. évf., 682-702.
- de Blij, H. (2012). *Why geography matters: More than ever*. Oxford University Press.
- Blumenthal, D., Campbell, E. G., Causino, N., & Louis, K. S. (1996). Participation of life-science faculty in research relationships with industry. *New England journal of medicine*, 335(23), 1734-1739. <https://doi.org/10.1056/NEJM199612053352305>
- Bollobás, B. (1998). Random graphs. In Uő, *Modern graph theory* (pp. 215-252). New York, (NY): Springer. https://doi.org/10.1007/978-1-4612-0619-4_7
- Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of mathematical sociology*, 2(1), 113-120. <https://doi.org/10.1080/0022250X.1972.9989806>
- Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, 92(5), 1170-1182. <https://doi.org/10.1086/228631>
- Borgatti, S. P., & Everett, M. G. (1999). Models of core/periphery structures. *Social networks*, 21(4), 375-395. [https://doi.org/10.1016/S0378-8733\(99\)00019-2](https://doi.org/10.1016/S0378-8733(99)00019-2)
- Borgatti, S. P., & Cross, R. (2003). A relational view of information seeking and learning in social networks. *Management Science*, 49(4), 432-445. <https://doi.org/10.1287/mnsc.49.4.432.14428>
- Bornmann, L., & Marx, W. (2013). How good is research really?: Measuring the citation impact of publications with percentiles increases correct assessments and fair comparisons. *EMBO reports*, 14(3), 226-230. <https://doi.org/10.1038/embor.2013.9>
- Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H. D. (2008). Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in science and environmental politics*, 8(1), 93-102. <https://doi.org/10.3354/esep00084>
- Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies*, 39(1), 61-74. <https://doi.org/10.1080/0034340052000320887>
- Boyack, K. (2008). Using detailed maps of science to identify potential collaborations. *Scientometrics*, 79(1), 27-44. <https://doi.org/10.1007/s11192-009-0402-6>

- Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?. *Journal of the Association for Information Science and Technology*, 61(12), 2389-2404. <https://doi.org/10.1002/asi.21419>
- Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351-374. <https://doi.org/10.1007/s11192-005-0255-6>
- Böde, C., Kovács, I. A., Szalay, M. S., Palotai, R., Korcsmáros, T., & Csermely, P. (2007). Network analysis of protein dynamics. *Febs Letters*, 581(15), 2776-2782. <https://doi.org/10.1016/j.febslet.2007.05.021>
- Börner, K., Maru, J. T., & Goldstone, R. L. (2004). The simultaneous evolution of author and paper networks. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5266-5273. <https://doi.org/10.1073/pnas.0307625100>
- Braam, R. R., Moed, H. F., & Van Raan, A. F. (1991a). Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for information science*, 42(4), 233-251. [https://doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4<233::AID-ASI1>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-4571(199105)42:4<233::AID-ASI1>3.0.CO;2-I)
- Braam, R. R., Moed, H. F., & Van Raan, A. F. (1991b). Mapping of science by combined co-citation and word analysis. II: Dynamical aspects. *Journal of the American Society for Information Science*, 42(4), 252-266. [https://doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4<252::AID-ASI2>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-4571(199105)42:4<252::AID-ASI2>3.0.CO;2-G)
- Braun, T., Gómez, I., Méndez, A., & Schubert, A. (1992). International co-authorship patterns in physics and its subfields, 1981-1985. *Scientometrics*, 24(2), 181-200. <https://doi.org/10.1007/BF02017908>
- Burris, V. (2005). Interlocking directorates and political cohesion among corporate elites. *American Journal of Sociology*, 111(1), 249-283. <https://doi.org/10.1086/428817>
- Cairncross, F. (1995). Telecommunications: the death of distance. *The Economist*, 30, 5-28.
- Cairncross, F. (2001). *The death of distance: 2.0: How the communications revolution will change our lives*. Texere.
- Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Information (International Social Science Council)*, 22(2), 191-235. <https://doi.org/10.1177/053901883022002003>
- Cameron, B. D. (2005). Trends in the usage of ISI bibliometric data: Uses, abuses, and implications. *portal: Libraries and the Academy*, 5(1), 105-125. <https://doi.org/10.1353/pla.2005.0003>
- Carley, S., Porter, A. L., Rafols, I., & Leydesdorff, L. (2017). Visualization of disciplinary profiles: Enhanced science overlay maps. *Journal of Data and Information Science*, 2(3), 68-111. <https://doi.org/10.1515/jdis-2017-0015>
- Chorley, R. J., & Haggett, P. (1969). *Network Analysis in Geography*. London: Edward Arnold.

- Christakis, N. A., & Fowler, J. H. (2010). *Kapcsolatok hálójában: Mire képesek a közösségi hálózatok, és hogyan alakítják sorsunkat?* Budapest: Typotex.
- Clarke, B. L. (1964). Multiple authorship trends in scientific papers. *Science*, 143(3608), 822-824. <https://doi.org/10.1126/science.143.3608.822>
- Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6), 066111. <https://doi.org/10.1103/PhysRevE.70.066111>
- Cowan, R. W. (2004). *Network models of innovation and knowledge diffusion*, MERIT Research Memorandum RM2004-016.
- Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago, University of Chicago Press.
- Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices?. *Journal of the American Society for Information Science and Technology*, 52(7), 558-569. <https://doi.org/10.1002/asi.1097>
- Cross, R., Borgatti, S. P., & Parker, A. (2001). Beyond answers: dimensions of the advice network. *Social Networks*, 23(3), 215-235. [https://doi.org/10.1016/S0378-8733\(01\)00041-7](https://doi.org/10.1016/S0378-8733(01)00041-7)
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1-9.
- Csermely, P. (2005). *A rejtett hálózatok ereje*. Budapest: Vince Kiadó.
- Csomós, Gy. (2017). Az európai városok tudományos kibocsátásának feltérképezése: egy területi tudománymetria elemzés a Scopus adatbázis alapján. *Területi Statisztika: a Központi Statisztikai Hivatal folyóirata*, 57(4), 356-387. <https://doi.org/10.15196/TS570402>
- Csomós, Gy. (2018). A spatial scientometric analysis of the publication output of cities worldwide. *Journal of Informetrics*, 12(2), 547-566. <https://doi.org/10.1016/j.joi.2018.05.003>
- Dance, A. (2012). Authorship: Who's on first?. *Nature*, 489(7417), 591-593. <https://doi.org/10.1038/nj7417-591a>
- Dietz, E. J. (1983). Permutation tests for association between two distance matrices. *Systematic Biology*, 32(1), 21-26. <https://doi.org/10.1093/sysbio/32.1.21>
- Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information processing & management*, 37(6), 817-842. [https://doi.org/10.1016/S0306-4573\(00\)00051-0](https://doi.org/10.1016/S0306-4573(00)00051-0)
- Doreian, P., Batagelj, V., & Ferligoj, A. (2004). Generalized blockmodeling of two-mode network data. *Social networks*, 26(1), 29-53. <https://doi.org/10.1016/j.socnet.2004.01.002>

- Dow, M. M., & Cheverud, J. M. (1985). Comparison of distance matrices in studies of population structure and genetic microdifferentiation: quadratic assignment. *American Journal of Physical Anthropology*, 68(3), 367-373. <https://doi.org/10.1002/ajpa.1330680307>
- Dusek, T., & Kotosz, B. (2016). *Területi statisztika*. Budapest, Akadémiai Kiadó. <https://doi.org/10.1556/9789634540014>
- Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1), 17-60.
- Everett, M. G., & Borgatti, S. P. (2014). Networks containing negative ties. *Social Networks*, 38, 111-120. <https://doi.org/10.1016/j.socnet.2014.03.005>
- Fábián Zs. (2012): A térparaméterek értelmezése és szerepük a területi kutatásokban. *Területi Statisztika*, 15, 177-215.
- Fan, K. W. (2015). Bias and other limitations affect measures of journals in integrative and complementary medicine. *Journal of the Medical Library Association: JMLA*, 103(3), 148. <https://doi.org/10.3163/1536-5050.103.3.009>
- Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41. <https://doi.org/10.2307/3033543>
- Freeman, L. C. (1979): Centrality in Networks: I. Conceptual clarification. *Social Networks*, 1, 215-239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Frenken, K., Hözl, W., & de Vor, F. (2005). The citation impact of research collaborations: the case of European biotechnology and applied microbiology (1988–2002). *Journal of Engineering and technology Management*, 22(1-2), 9-30. <https://doi.org/10.1016/j.jengtecman.2004.11.002>
- Frenken, K., Hardeman, S., & Hoekman, J. (2009a). Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics*, 3(3), 222-232. <https://doi.org/10.1016/j.joi.2009.03.005>
- Frenken, K., Hoekman, J., Kok, S., Ponds, R., van Oort, F., & van Vliet, J. (2009b). Death of distance in science? A gravity approach to research collaboration. In A. Pyka, & A. Scharnhorst (szerk.), *Innovation networks* (pp. 43-57). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-92267-4_3
- Fritsch, A. (2012). *mcclust: Process an MCMC Sample of Clusterings*. R package version 1.0. <https://CRAN.R-project.org/package=mcclust>
- Garfield, E. (1971). The mystery of the transposed journal lists—wherein Bradford's Law of Scattering is generalized according to Garfield's Law of Concentration. *Current Contents*, 3(33), 5-6.
- Garfield, E. (1988). Announcing the SCI compact disk edition: CD-ROM gigabyte storage technology, novel software, and bibliographic coupling make desk-top research and discovery a reality. *Current Contents*, (22), 3-13.
- Gibbons, M. (Ed.). (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. Sage.

- Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12), 7821-7826. <https://doi.org/10.1073/pnas.122653799>
- Glänzel, W. (2002). Co-authorship patterns and trends in the sciences (1980–1998). A bibliometric study with implications for database indexing and search strategies, *Library Trends*, 50, 461–473.
- Glänzel, W., & Czerwon, H. J. (1996). A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics*, 37(2), 195-221. <https://doi.org/10.1007/BF02093621>
- Glänzel, W., & Schubert, A. (2004). Analysing scientific networks through co-authorship. In H. F. Moed, W. Glänzel, & U. Schmoch (szerk.), *Handbook of quantitative science and technology research* (pp. 257-276). Dordrecht: Springer. https://doi.org/10.1007/1-4020-2755-9_12
- Goldfinch, S., Dale, T., & DeRouen, K. (2003). Science from the periphery: Collaboration, networks and 'Periphery Effects' in the citation of New Zealand Crown Research Institutes articles, 1995-2000. *Scientometrics*, 57(3), 321-337. <https://doi.org/10.1023/A:1025048516769>
- Gordon, M. (1980). A critical reassessment of inferred relations between multiple authorship, scientific collaboration, the production of papers and their acceptance for publication. *Scientometrics*, 2(3), 193-201. <https://doi.org/10.1007/BF02016697>
- Granovetter M. (1973): The strength of weak ties. *American Journal of Sociology*, 78,(6), 1360-1380. <https://doi.org/10.1086/225469>
- Granovetter M. S. (1983): The strength of the weak ties: A network theory revisited. *Sociological Theory*, 1, 201-233. <https://doi.org/10.2307/202051>
- Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American journal of sociology*, 91(3), 481-510. <https://doi.org/10.1086/228311>
- Grossetti, M., Eckert, D., Gingras, Y., Jégou, L., Larivière, V., & Milard, B. (2014). Cities and the geographical deconcentration of scientific activity: A multilevel analysis of publications (1987–2007). *Urban Studies*, 51(10), 2219-2234. <https://doi.org/10.1177/0042098013506047>
- Hall, B. H. (2003). On copyright and patent protection for software and databases: a tale of two worlds. In O. Granstrand (szerk.), *Economics, Law and Intellectual Property* (pp. 259-277). Boston (MA), Springer. https://doi.org/10.1007/978-1-4757-3750-9_12
- Hamers, L., Hemeryck, Y., Herweyers, G., Janssen, M., Keters, H., Rousseau, R., & Vanhoutte, A. (1989). Similarity measures in scientometric research: the Jaccard index versus Salton's cosine formula. *Information Processing & Management*, 25(3), 315-318. [https://doi.org/10.1016/0306-4573\(89\)90048-4](https://doi.org/10.1016/0306-4573(89)90048-4)
- Heimeriks, G., Hoerlesberger, M., & Van den Besselaar, P. (2003). Mapping communication and collaboration in heterogeneous research networks. *Scientometrics*, 58(2), 391-413. <https://doi.org/10.1023/A:1026296812830>

- Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, 44(2), 193–215. <https://doi.org/10.1007/BF02457380>
- Hicks, D. (2004). The four literatures of social science. In H. F. Moed, W. Glänzel, & U. Schmoch (szerk.), *Handbook of quantitative science and technology research* (pp. 473–496). Springer, Dordrecht. https://doi.org/10.1007/1-4020-2755-9_22
- Hoekman, J., Frenken, K., & Van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science*, 43(3), 721–738. <https://doi.org/10.1007/s00168-008-0252-9>
- Hoekman, J., Frenken, K., & Tijssen, R. J. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, 39(5), 662–673. <https://doi.org/10.1016/j.respol.2010.01.012>
- Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of scientific collaboration networks in Scientometrics. *Scientometrics*, 75(2), 189–202. <https://doi.org/10.1007/s11192-007-1771-3>
- Hubert, L. (1986). *Assignment methods in combinational data analysis* (Vol. 73). CRC Press.
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2(1), 193–218. <https://doi.org/10.1007/BF01908075>
- Hubert, L., & Schultz, J. (1976). Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 29(2), 190–241. <https://doi.org/10.1111/j.2044-8317.1976.tb00714.x>
- Hudson, J. (1996). Trends in multi-authored papers in economics. *The Journal of Economic Perspectives*, 10(3), 153–158. <https://doi.org/10.1257/jep.10.3.153>
- Ikpaaahindi, L. [Schubert, A. ford.] (1986). An overview of bibliometrics: its measurements, laws and their applications. [A bibliometria és a tudománymetria mérési módszerei, törvényei és alkalmazásai]. *Kutatás-fejlesztés, Tudományos-szervezési tájékoztató*. 26(3–4), 279–294.
- Jacso, P. (2005). As we may search—Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537.
- Jakobi, Á. (2016). Az információs világ megújuló területi kutatása a big data korszakban: lehetőségek és tapasztalatok. Habilitációs értekezés, Eötvös Loránd Tudományegyetem, Budapest.
- Jakobi Á., & Jeney L. (2008). A szomszédsági mátrix problematikájának megoldási lehetőségei – elmélet és gyakorlat. *Területi Statisztika*, 3, 295–304.
- Janssens, F., Leta, J., Glänzel, W., & De Moor, B. (2006). Towards mapping library and information science. *Information processing & management*, 42(6), 1614–1642. <https://doi.org/10.1016/j.ipm.2006.03.025>

- Janssens, F., Zhang, L., De Moor, B., & Glänzel, W. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing & Management*, 45(6), 683-702. <https://doi.org/10.1016/j.ipm.2009.06.003>
- Jarneving, B. (2007). Bibliographic coupling and its application to research-front and other core documents. *Journal of Informetrics*, 1(4), 287-307. <https://doi.org/10.1016/j.joi.2007.07.004>
- Katz, J. S. (1992). Bibliometric assessment of intranational university-university collaboration. PhD értekezés, University of Sussex, Nagy-Britannia.
- Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31-43. <https://doi.org/10.1007/BF02018100>
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration?. *Research policy*, 26(1), 1-18. [https://doi.org/10.1016/S0048-7333\(96\)00917-1](https://doi.org/10.1016/S0048-7333(96)00917-1)
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *Journal of the Association for Information Science and Technology*, 14(1), 10-25. <https://doi.org/10.1002/asi.5090140103>
- Kessler, M. M. (1965). Comparison of the results of bibliographic coupling and analytic subject indexing. *American Documentation*, 16(3), 223-233. <https://doi.org/10.1002/asi.5090160309>
- Krackhardt, D. (1987). QAP partialling as a test of spuriousness. *Social networks*, 9(2), 171-186. [https://doi.org/10.1016/0378-8733\(87\)90012-8](https://doi.org/10.1016/0378-8733(87)90012-8)
- Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social networks*, 10(4), 359-381. [https://doi.org/10.1016/0378-8733\(88\)90004-4](https://doi.org/10.1016/0378-8733(88)90004-4)
- Kraut, R., Egidio, C., & Galegher, J. (1988). Patterns of contact and communication in scientific research collaboration. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work* (pp. 1-12). ACM. <https://doi.org/10.1145/62266.62267>
- Kretschmer, H. (1994). Coauthorship networks of invisible colleges and institutionalized communities. *Scientometrics*, 30(1), 363-369. <https://doi.org/10.1007/BF02017234>
- Kürtösi, Zs. (2004). A társadalmi kapcsolatháló elemzés módszertani alapjai. In L. Leteyei (szerk.), *Településkutatás* (pp. 663-684). Budapest: L'Harmattan.
- LaFollette, M. C. (1992). *Stealing into print: fraud, plagiarism, and misconduct in scientific publishing*. Univ of California Press.
- Laki, J., & Palló, G. (2001). Projektvilág és informális hálózat a tudományban. In: K. Nyíri, (szerk.), *A 21. századi kommunikáció új útjai* (pp. 173-193.). Budapest: MTA Filozófiai Kutatóintézet.
- Langfelder, P., Zhang, B., & Horvath S. (2016). *dynamicTreeCut: Methods for Detection of Clusters in Hierarchical Clustering Dendrograms. R package version 1.63-1*. <https://CRAN.R-project.org/package=dynamicTreeCut>

- Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, 57(8), 997-1004. <https://doi.org/10.1002/asi.20349>
- Laudel, G. (2001). Collaboration, creativity and rewards: why and how scientists collaborate. *International Journal of Technology Management*, 22(7-8), 762-781. <https://doi.org/10.1504/IJTM.2001.002990>
- Lee, J. S., Cho, H., Gay, G., Davidson, B., & Ingraffea, A. R. (2003). Technology acceptance and social networking in distance learning. *Educational Technology & Society*, 6(2), 50-61.
- Van Leeuwen, T., Moed, H., Tijssen, R., Visser, M., & Van Raan, A. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics*, 51(1), 335-346. <https://doi.org/10.1023/A:1010549719484>
- Legendre, P. (2000). Comparison of permutation methods for the partial correlation and partial Mantel tests. *Journal of Statistical Computation and Simulation*, 67(1), 37-73. <https://doi.org/10.1080/00949650008812035>
- Lengyel, B. (2005) Triple Helix kapcsolatok a tudásmenedzsment szemszögéből. In N. Buzás (szerk.), *Tudásmenedzsment és tudásalapú gazdaságfejlesztés. SZTE Gazdaságtudományi Kar Közleményei* (pp. 293-311). Szeged: JATE Press.
- Lengyel, B., Varga, A., Ságvári, B., & Jakobi, Á. (2013). Distance dead or alive: online social networks from a geography perspective. *IBS Working Papers*, 1/2013. https://www.ibs-b.hu/data/downloads/2015/09/10/OSON_WP2_iWiWdistance.pdf
- Lengyel, B., Varga, A., Ságvári, B., Jakobi, Á., & Kertész, J. (2015). Geographies of an online social network. *PloS one*, 10(9), e0137248. <https://doi.org/10.1371/journal.pone.0137248>
- Lengyel, B., Varga, A., Jakobi, Á., Kertész, J., & Ságvári, B. (2016). Az iWiW földrajza. *Területi Statisztika: a Központi Statisztikai Hivatal folyóirata*, 56(1), 30-45. <https://doi.org/10.15196/TS560103>
- Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, 58(9), 1303-1319. <https://doi.org/10.1002/asi.20614>
- Leydesdorff, L., & Etzkowitz, H. (1998). The triple helix as a model for innovation studies. *Science and Public Policy*, 25(3), 195-203.
- Leydesdorff, L., & Persson, O. (2010). Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes. *Journal of the American Society for Information Science and Technology*, 61(8), 1622-1634. <https://doi.org/10.1002/asi.21347>
- Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, 60(2), 348-362. <https://doi.org/10.1002/asi.20967>

- Leydesdorff, L., Carley, S., & Rafols, I. (2013). Global maps of science based on the new Web-of-Science categories. *Scientometrics*, 94(2), 589-593. <https://doi.org/10.1007/s11192-012-0784-8>
- Li, M. X., Palchykov, V., Jiang, Z. Q., Kaski, K., Kertész, J., Micciché, S., ... & Mantegna, R. N. (2014). Statistically validated mobile communication networks: the evolution of motifs in European and Chinese data. *New Journal of Physics*, 16(8), 083038. <https://doi.org/10.1088/1367-2630/16/8/083038>
- Liang, L., Kretschmer, H., Guo, Y. & Beaver, D. D. (2000). Age structure of two-dimensional and three-dimensional research collaboration in Chinese computer science. *Collaboration in Science and in Technology*, 143.
- Liang, L.M., Zhang, L., Kretschmer, H., Scharnhorst, A., 2006. Geographical and linguistic preferences in scientific collaboration of the European Union (1994–2003). In: *Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting, Nancy, France*.
- Liang, L., Rousseau, R., & Zhong, Z. (2013). Non-English journals and papers in physics and chemistry: bias in citations?. *Scientometrics*, 95(1), 333-350. <https://doi.org/10.1007/s11192-012-0828-0>
- Lievrouw, L. A., Rogers, E. M., Lowe, C. U., & Nadel, E. (1987). Triangulation as a research strategy for identifying invisible colleges among biomedical scientists. *Social Networks*, 9(3), 217-248. [https://doi.org/10.1016/0378-8733\(87\)90021-9](https://doi.org/10.1016/0378-8733(87)90021-9)
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington academy of sciences*, 16(12), 317-323.
- Lőcsei, H., & Szalkai, G. (2008). Helyzeti és fejlettségi centrum–periféria relációk a hazai kistérségekben. *Területi Statisztika*, 48(3), 305-314.
- Luce, R. D., & Perry, A. D. (1949). A method of matrix analysis of group structure. *Psychometrika*, 14(2), 95-116. <https://doi.org/10.1007/BF02289146>
- Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology, & Human Values*, 17(1), 101-126. <https://doi.org/10.1177/016224399201700106>
- Luukkonen, T., Tijssen, R., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15-36. <https://doi.org/10.1007/BF02016282>
- Macmillan English dictionary for advanced learners* (2005). Oxford, Macmillan.
- Maglaughlin, K. L., & Sonnenwald, D. H. (2005). Factors that impact interdisciplinary natural science research collaboration in academia. In *Proceedings of the Conference of the International Society for Scientometrics and Informetrics*, 499–508.
- Magyar értelmező kéziszótár* (1972). Budapest, Akadémiai Kiadó.

- Mali, F., Kronegger, L., Doreian, P., & Ferligoj, A. (2012). Dynamic scientific co-authorship networks. In A. Scharnhorst, K. Börner, & P. van den Besselaar (szerk.), *Models of science dynamics* (pp. 195-232). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-23068-4_6
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1), 209-220.
- Martin, B. R., Irvine, J., Narin, F., Stevens, K. A., & Sterritt, C. (1990). Recent trends in the output and impact of British science. *Science and Public Policy*, 17(1), 14-26.
- Martyn, J. (1964). Bibliographic coupling. *Journal of Documentation*, 20(4), 236-236. <https://doi.org/10.1108/eb026352>
- Mattsson, P., Laget, P., Nilsson, A., & Sundberg, C. J. (2008). Intra-EU vs. extra-EU scientific co-publication patterns in EU. *Scientometrics*, 75(3), 555-574. <https://doi.org/10.1007/s11192-007-1793-x>
- McCain, K. W., Verner, J. M., Hislop, G. W., Evanco, W., & Cole, V. (2005). The use of bibliometric and knowledge elicitation techniques to map a knowledge domain: Software engineering in the 1990s. *Scientometrics*, 65(1), 131-144. <https://doi.org/10.1007/s11192-005-0262-7>
- Meadows, A. J., & O'Connor, J. G. (1971). Bibliographical statistics as a guide to growth points in science. *Science Studies*, 1(1), 95-99. <https://doi.org/10.1177/030631277100100107>
- Melin, G. (2000). Pragmatism and self-organization: Research collaboration on the individual level. *Research policy*, 29(1), 31-40. [https://doi.org/10.1016/S0048-7333\(99\)00031-1](https://doi.org/10.1016/S0048-7333(99)00031-1)
- Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, 36(3), 363-377. <https://doi.org/10.1007/BF02129600>
- Meyer, D., Buchta, C. (2015). *proxy: Distance and Similarity Measures*. R package version 0.4-15. <https://CRAN.R-project.org/package=proxy>
- Milgram, S. (1967). The Small World Problem. *Psychology Today*, 2, 60-67. <https://doi.org/10.1037/e400002009-005>
- Moed, H., De Bruin, R., Nederhof, A., & Tijssen, R. (1991). International scientific co-operation and awareness within the European Community: Problems and perspectives. *Scientometrics*, 21(3), 291-311. <https://doi.org/10.1007/BF02093972>
- Mongeon, P., & Paul-Hus, A. (2014). The journal coverage of bibliometric databases: A comparison of Scopus and Web of Science. *Metrics Seattle. Full-text doi*, 10(2.1), 4759-7762.
- Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69(2), 213-238. <https://doi.org/10.1177/000312240406900204>
- Moreno, J. L. (1934). *Who shall survive?* Beacon (NY): Beacon House.

- MTA KIK TTO (2013). Az MTA tudományos teljesítményének szerkezeti és minőségi jellemzői a 2013-as évben (Tudományometriai elemzés a Web of Science és az MTMT integrált adatai alapján). <http://www.mtakszi.iif.hu/jelentesek.php>
- MTA KIK TTO (2014). Az MTA-kutatóhálózat tudományos teljesítményének szerkezeti jellemzői 2014. <http://www.mtakszi.iif.hu/jelentesek.php>
- MTA KIK TTO (2015). Az MTA-kutatóhálózat tudományos teljesítményének szerkezeti jellemzői 2015. <http://www.mtakszi.iif.hu/jelentesek.php>
- MTA KIK TTO (2016). Az MTA-kutatóhálózat tudományos teljesítményének szerkezeti jellemzői 2016. <http://www.mtakszi.iif.hu/jelentesek.php>
- MTA KIK TTO (2017). Az MTA-kutatóhálózat tudományos teljesítményének szerkezeti jellemzői 2015-2017. <http://www.mtakszi.iif.hu/jelentesek.php>
- Nagy, A. M. (2016). Nemzetközi publikációs verseny a közgazdaságtudományban: Módszertani javaslatok a tudománymetria területéről. PhD disszertáció, Pannon Egyetem, Veszprém.
- Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill (NJ): Computer Horizons.
- Narin, F., Stevens, K., & Whitlow, E. (1991). Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics*, 21(3), 313-323 <https://doi.org/10.1007/BF02093973>
- Nemes Nagy, J. (1998). *A tér a társadalomkutatásban*. Budapest: Hilscher Rezső Szociálpolitikai Egyesület.
- Nemes Nagy, J. (2003). A regionális tudomány dualitása és paradigmái–hazai tükrök. *Tér és Társadalom*, 17, 1-17. <https://doi.org/10.17649/TET.17.1.871>
- Nemes Nagy, J. (2007). Kvantitatív társadalmi térelemzési eszközök a mai regionális tudományban. *Tér és Társadalom*, 21(1), 1-19. <https://doi.org/10.17649/TET.21.1.1090>
- Nemes Nagy, J. (2009). *Terek, helyek, régiók. A regionális tudomány alapjai*. Budapest: Akadémiai Kiadó.
- Nemes Nagy, J. (2017). Tér, függés, kohézió, hálózatok. *Területi Statisztika*, 57(1), 3-23.
- Newman, M. E. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical review E*, 64(1), 016131. <https://doi.org/10.1103/PhysRevE.64.016131>
- Newman, M. E. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical review E*, 64(1), 016132. <https://doi.org/10.1103/PhysRevE.64.016132>
- Newman, M. E. (2001c). The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, 98(2), 404-409. <https://doi.org/10.1073/pnas.98.2.404>

- Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2), 167-256. <https://doi.org/10.1137/S003614450342480>
- Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5200-5205. <https://doi.org/10.1073/pnas.0307545100>
- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), 8577-8582. <https://doi.org/10.1073/pnas.0601602103>
- Nomaler, Z. Ö., Frenken, K. K., & Heimeriks, G. (2013). Do more distant collaborations have more citation impact?. *Journal of Informetrics*, 7(4), 966. <https://doi.org/10.1016/j.joi.2013.10.001>
- Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., & Van den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research policy*, 36(7), 1016-1034. <https://doi.org/10.1016/j.respol.2007.04.003>
- O'Brien, R. (1992). *Global financial integration: The end of geography*. Royal Institute of International Affairs.
- Paier, M., & Scherngell, T. (2011). Determinants of collaboration in European R&D networks: empirical evidence from a discrete choice model. *Industry and Innovation*, 18(1), 89-104. <https://doi.org/10.1080/13662716.2010.528935>
- Patel, N. (1973). Collaboration in the professional growth of American sociology. *Information (International Social Science Council)*, 12(6), 77-92.
- Pénzes, J. (2013): The dimension of peripheral areas and their restructuring in Central Europe. *Hungarian Geographical Bulletin* 62(4), 373-386.
- Pénzes, J., & Hegedűs, L. (2016): A területfejlesztés kedvezményezett településeinek vizsgálata térinformatikai eszközökkel. In B. Balázs (szerk.), *Az elmélet és a gyakorlat találkozása a térinformatikában VII.* (pp. 381-386). Debrecen: Debreceni Egyetemi Kiadó.
- Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421-432. <https://doi.org/10.1023/B:SCIE.0000034384.35498.7d>
- Ponds, R. (2009). The limits to internationalization of scientific research collaboration. *The Journal of Technology Transfer*, 34(1), 76-94. <https://doi.org/10.1007/s10961-008-9083-1>
- Ponds, R., Van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science*, 86(3), 423-443. <https://doi.org/10.1111/j.1435-5957.2007.00126.x>
- de Solla Price, D. (1979): *Kis tudomány – nagy tudomány*. Budapest: Akadémiai Kiadó.
- de Solla Price, D. J., & Beaver, D. (1966). Collaboration in an invisible college. *American Psychologist*, 21(11), 1011-1018. <https://doi.org/10.1037/h0024051>

- Probáld, F. (2007). Társadalomföldrajz és regionális tudomány. *Tér és Társadalom*, 21(1), 21-33. <https://doi.org/10.17649/TET.21.1.1091>
- R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Radicchi, F., Fortunato, S., Markines, B., & Vespignani, A. (2009). Diffusion of scientific credits and the ranking of scientists. *Physical Review E*, 80(5), 056103. <https://doi.org/10.1103/PhysRevE.80.056103>
- Rafols, I. (2017. augusztus). *Questioning the agenda of indicators of societal impact*. Plenárius előadás a Science, Technology & Information Indicators konferencián, Párizs, Franciaország.
- Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the Association for Information Science and Technology*, 61(9), 1871-1887. <https://doi.org/10.1002/asi.21368>
- Rahman, A. I. M., Guns, R., Rousseau, R., & Engels, T. C. (2017). Cognitive distances between evaluators and evaluatees in research evaluation: A comparison between three informetric methods at the journal and subject category aggregation level. *Frontiers in Research Metrics and Analytics*, 2, 6. <https://doi.org/10.3389/frma.2017.00006>
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336), 846-850. <https://doi.org/10.1080/01621459.1971.10482356>
- Rechnitzer, J. (2007). A társadalomtudomány új ága: a regionális tudomány. *Magyar Tudomány*, 168(17), 1580-1590.
- Robins, G. (2013). A tutorial on methods for the modeling and analysis of social network data. *Journal of Mathematical Psychology*, 57(6), 261-274. <https://doi.org/10.1016/j.jmp.2013.02.001>
- Rollié, S., Mangold, M., & Sundmacher, K. (2012). Designing biological systems: systems engineering meets synthetic biology. *Chemical Engineering Science*, 69(1), 1-29. <https://doi.org/10.1016/j.ces.2011.10.068>
- Salager-Meyer, F., Alcaraz Ariza, M. Á., Pabón, M., & Zambrano, N. (2006). Paying one's intellectual debt: Acknowledgments in scientific/conventional and complementary/alternative medical research. *Advances in medical discourse analysis: Oral and written contexts*, 407-430.
- Salton, G., McGill, M.J. (1983). *Introduction to modern information retrieval*. Auckland, McGraw-Hill
- Schott, T. (1998). Ties between center and periphery in the scientific world-system: accumulation of rewards, dominance and self-reliance in the center. *Journal of World-Systems Research*, 4(2), 112-144. <https://doi.org/10.5195/jwsr.1998.148>
- Schubert, A., & Braun, T. (1990). International collaboration in the sciences 1981-1985. *Scientometrics*, 19(1-2), 3-10. <https://doi.org/10.1007/BF02130461>

Scopus Content Coverage Guide, 2017 August, Elsevier

Scott, J. (2000). *Social network analysis: a handbook*. London, Sage Publication.

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using gaussian finite mixture models. *The R journal*, 8(1), 289. <https://doi.org/10.32614/RJ-2016-021>

Siguan, M. (2001). English and the Language of Science: On the Unity. In U. Ammon (szerk.), *The dominance of English as a language of science: Effects on other languages and language communities* (pp. 59-71). Berlin, Mouton de Gruyter.

Simonovits A. (2009). Válogatott fejezetek a matematika történetéből. Budapest: Typotex Kiadó.

Sivertsen, G. (2009). Publication patterns in all fields. In F. Åström, R. danell, B. Larsen, & J.W. Schneider (szerk.), *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th birthday* (pp. 55-60.). International Society for Scientometrics and Informetrics. <http://portal.research.lu.se/portal/files/5902071/1458992.pdf>

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the Association for Information Science and Technology*, 24(4), 265-269. <https://doi.org/10.1002/asi.4630240406>

Small, H., & Garfield, E. (1985). The geography of science: disciplinary and national mappings. *Journal of Information Science*, 11(4), 147-159. <https://doi.org/10.1177/016555158501100402>

Smith, M. (1958). The trend toward multiple authorship in psychology. *American Psychologist*, 13(10), 596-599. <https://doi.org/10.1037/h0040487>

Sonnenwald, D.H. (2007). Scientific collaboration. In B. Cronin (szerk.), *Annual Review of Information Science and Technology*, Vol. 41 (pp. 643-681), Medford (NJ): Information Today. <https://doi.org/10.1002/aris.2007.1440410121>

Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review. *Information Scientist*, 6(1), 33-38. <https://doi.org/10.1177/016555158300600105>

Tagai G. (2012): Térkapcsolati modellek a társadalmi térben értelmezett fekvés vizsgálatában. In: Bottlik Zs., Czirfusz M., Gyapay B., Kőszegi M., & Pfening V. (szerk.), *Társadalomföldrajz-Területfejlesztés-Regionális tudomány* (pp. 177-191.). Budapest: ELTE TTK Földtudományi Doktori Iskola.

Tagliacozzo, R. (1967). Citations and citation indexes: A review. *Methods of Information in Medicine*, 6(03), 136-142. <https://doi.org/10.1055/s-0038-1636363>

Tan, P. N., Steinbach, M., Karpatne, A., & Kumar, V. (2007). *Introduction to data mining*. Pearson Education Inc.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008, August). Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 990-998). ACM. <https://doi.org/10.1145/1401890.1402008>

- Teleki, P. (1917/1996). *A földrajzi gondolat története*. Budapest: Kossuth Kiadó.
- Testa, J. (2016). Web of science core collection: Journal selection process. *Wok Info/ Web of Science*.
- Tobler, W. R. (1970): Computer Movie Simulating Urban Groth in the Detroit Region. *Economic Geography*, 46, 2., 234-240. <https://doi.org/10.2307/143141>
- Torre, A. & Gilly, J. P. (2000). On the analytical dimension of proximity dynamics. *Regional studies*, 34(2), 169-180. <https://doi.org/10.1080/00343400050006087>
- Tóth G. (2006): Centrum-periféria viszonyok vizsgálata a hazai közúthálózaton. *Területi statisztika* 9(5), 476-493.
- Tóth G., Kincses Á. (2007): Közúti elérhetőségi vizsgálatok Európában. *Statisztikai Szemle*, 85(5), 432-463.
- Tóth Zs. (2009). *Tudáshálók a gazdasági felsőoktatásban*. PhD disszertáció, Nyugat-magyarországi Egyetem, Sopron.
- Varga, A. V. (2011). Measuring the semantic integrity of scientific fields: a method and a study of sociology, economics and biophysics. *Scientometrics*, 88(1), 163-177. <https://doi.org/10.1007/s11192-011-0342-9>
- Vicsek, L. (2006). *Fókuszcsoport: elméleti megfontolások és gyakorlati alkalmazás*. Budapest, Osiris Kiadó.
- Vida, Zs. (2013). A térkapcsolatok értelmezési lehetőségei hálózatokon In: Bottlik Zs. (szerk.), *Önálló lépések a tudomány területén* (pp. 115–129.). Budapest: ELTE Földtudományi Doktori Iskola.
- Vida, Zs. V. (2016). Scientific collaborations at the level of countries—a case study: A “hard” Physical geography and a “soft” Economics science field. *DETUROPE*, 8(3), 224-246.
- Vieira, E. S., & Gomes, J. A. N. F. (2009). A comparison of Scopus and Web of Science for a typical university. *Scientometrics*, 81(2), 587-600. <https://doi.org/10.1007/s11192-009-2178-0>
- Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, 34(10), 1608-1618. <https://doi.org/10.1016/j.respol.2005.08.002>
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511815478>
- Watts, D. J. (2003) *Six degrees. The science of a connected age*. New York: W. W. Norton & Company.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684), 440. <https://doi.org/10.1038/30918>

- White, H. D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure?: Longitudinal evidence from the "Globenet" interdisciplinary research group. *Journal of the American Society for Information Science and Technology*, 55(2), 111-126. <https://doi.org/10.1002/asi.10369>
- Wuyts, S., Colombo, M. G., Dutta, S., & Nooteboom, B. (2005). Empirical tests of optimal cognitive distance. *Journal of Economic Behavior & Organization*, 58(2), 277-302. <https://doi.org/10.1016/j.jebo.2004.03.019>
- Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, co-authorship networks, and cword networks relate to each other. *Journal of the American Society for Information Science and Technology*, 63(7), 1313-1326. <https://doi.org/10.1002/asi.22680>
- Zhao, D., & Strotmann, A. (2008a). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic coupling analysis. *Journal of the American Society for Information Science and Technology*, 59(13), 2070-2086. <https://doi.org/10.1002/asi.20910>
- Zhao, D., & Strotmann, A. (2008b). Author bibliographic coupling: Another approach to citation-based author knowledge network analysis. *Proceedings of the Association for Information Science and Technology*, 45(1), 1-10. <https://doi.org/10.1002/meet.2008.1450450292>
- Ziman, J., (1994). *Prometheus Bound. Science in a Dynamic Steady State*. Cambridge, Cambridge Univ. Press. <https://doi.org/10.1017/CBO9780511585067>
- Zipf G. K. (1949): *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge (MA): Addison – Wesley.
- Zitt, M., Bassecoulard, E., & Okubo, Y. (2000). Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics*, 47(3), 627-657. <https://doi.org/10.1023/A:1005632319799>

Internetes források:

apps.webofknowledge.com. Megtekintve: 2018. 05. 25.

www.clarivate.com. Megtekintve: 2018. 05. 25.

www.cost.eu. Megtekintve: 2018. 05. 25.

www.doogal.co.uk/BatchGeocoding.php. Megtekintve: 2018. 05. 25.

ipscience-help.thomsonreuters.com. Megtekintve: 2018. 05. 25.

www.kfki.hu/~cheminfo/hun/eloado/adatb/adatb2.html. Megtekintve: 2018. 05. 25.

maps.google.com. Megtekintve: 2018. 05. 25.

www.news.clarivate.com. Megtekintve: 2018. 05. 25.

www.open.ac.uk/blogs/the_orb/?p=2062. Megtekintve: 2018. 05. 25.

www.r-bloggers.com/summary-of-community-detection-algorithms-in-igraph-0-6/. Megtekintve: 2018. 05. 25.

www.rkk.hu/hu. Megtekintve: 2018. 05. 25.

service.elsevier.com. Megtekintve: 2018. 05. 25.

wokinfo.com. Megtekintve: 2018. 05. 25.

Ábrák jegyzéke

1. ábra: Az értekezés felépítése	15
2. ábra: A königsbergi hidak gráfként való leképezése	18
3. ábra: A és C csúcs közötti út	19
4. ábra: Szabályos, Watts–Strogatz-féle kisvilág-, valamint véletlen modell	22
5. ábra: A Watts–Strogatz-modell tulajdonságai	22
6. ábra: Erdős–Rényi-féle véletlen gráf és fokszámeloszlása	24
7. ábra: Barabási–Albert-modellel létrehozott skálafüggetlen hálózat és fokszámeloszlása	24
8. ábra: A centrum-periféria duál diszkrét modellje: C a centrum, P a periféria jele, a cellákban lévő 1 érték a kapcsolat létére, a 0 annak hiányára utal	52
9. ábra: Kutatói együttműködési hálózatok vizsgálatának módszerei	63
10. ábra: Bibliográfiai csatolás és együtt hivatkozás: a körök publikációkat jelölnek, a fekete nyilak a hivatkozás irányát. Az A és B között fennálló kapcsolatot jelöli az átlátszó kétirányú nyíl	63
11. ábra: Bibliográfiai csatolás szerzői szintre vetítése P: publikáció, A: szerző, R: hivatkozás	73
12. ábra: A WoS és a Scopus tartalmának viszonya egymáshoz képest 2013-ban	80
13. ábra: A természetföldrajz (Geography, Physical) és a közgazdaságtan (Economics) elhelyezkedése egy 2015-ös tudománytérképen. A térkép a WoS 227 tudományterületi kategóriája alapján 18 klasztert különít el	83
14. ábra: Nyelvi lefedettség az SCI-ben 1980 és 1998 között	89
15. ábra: A három fő tudományterületi ág rekordjainak megoszlása a WoS-ban	90
16. ábra: Példa a WoS szerzői affiliációkat tartalmazó C1 mező tartalmára	95
17. ábra: A mintában szereplő szerzők megoszlása országok szerint a közgazdaságtan területén	104
18. ábra: A mintában szereplő szerzők megoszlása országok szerint a természetföldrajz területén	105
19. ábra: A 7. fejezet felépítése	109
20. ábra: Teljes kognitív (a,d), társadalmi (b,e) és tiszta kognitív (c,f) távolságmátrixok dendrogramjai a közgazdaságtan (a,b,c) és a természetföldrajz (d,e,f) területén	112
21. ábra: Klasztercsoportok méretének eloszlása a társadalmi távolságmátrixon a közgazdaságtan területén	113

22. ábra: Teljes kognitív (a,d), társadalmi (b,e) és tiszta kognitív (c,f) hálózatok komponensei a közgazdaságtan (a,b,c) és a természetföldrajz (d,e,f) területén	121
23. ábra: Teljes kognitív (a,d), társadalmi (b,e) és tiszta kognitív (c,f) hálózatok foksámeloszlása a közgazdaságtan (a,b,c) és a természetföldrajz (d,e,f) területén	123
24. ábra: A modularitás sematikus példája irányítatlan hálózatokon	129
25. ábra: Teljes kognitív (a,d), társadalmi (b) és tiszta kognitív (c,e) hálózatok fast greedy algoritmus alapján létrehozott moduljai (a,b,c) és óriáskomponenseik (d,e) a közgazdaságtan területén	130
26. ábra: Teljes kognitív (a,d), társadalmi (b,e) és tiszta kognitív (c,f) hálózatok fast greedy algoritmus alapján létrehozott moduljai (a,b,c) és óriáskomponenseik (d,e,f) a természetföldrajz területén	131
27. ábra: Modulok méretének eloszlása a teljes kognitív hálózaton a közgazdaságtan területén	13
28. ábra: A társadalmi, a teljes és a tiszta kognitív hálózatok a földrajzi térben a közgazdaságtan területén	137
29. ábra: A társadalmi, a teljes és a tiszta kognitív hálózatok a földrajzi térben a természetföldrajz területén	137
30. ábra: A társadalmi és a tiszta kognitív hálózatok medián feletti élsúllyal rendelkező kapcsolatai a földrajzi térben a természetföldrajz területén	139
31. ábra: A társadalmi és a tiszta kognitív hálózatok medián feletti élsúllyal rendelkező kapcsolatai Európában a közgazdaságtan területén	140
32. ábra: A társadalmi hálózat medián feletti élsúllyal rendelkező kapcsolatai Európában a természetföldrajz területén	140
33. ábra: A tiszta kognitív hálózat medián feletti élsúllyal rendelkező kapcsolatai Európában a természetföldrajz területén	141
34. ábra: A tiszta kognitív hálózat kapcsolatainak legerősebb 10%-a Európában a természetföldrajz területén	141
35. ábra: Magyarország és szomszédai közötti teljes kognitív kapcsolatok a közgazdaságtan területén	143
36. ábra: Magyarország és szomszédai közötti társadalmi és tiszta kognitív kapcsolatok a közgazdaságtan területén	143
37. ábra: Magyarország és szomszédai közötti medián feletti élsúllyal rendelkező társadalmi és tiszta kognitív kapcsolatok a természetföldrajz területén	143

Táblázatok jegyzéke

1. táblázat: Térkategóriák értelmezése hálózatokon	54
2. táblázat: WoS rekord felhasznált mezői (példa)	85
3. táblázat: Az adatbázis mérete – magyar affiliációval is rendelkező cikkek 2010 és 2014 között a két tudományterületen	86
4. táblázat: szerző-publikáció mátrix	98
5. táblázat: nyers társszerzői mátrix	98
6. táblázat: Salton-hasonlóságot tartalmazó társszerzői tábla	99
7. táblázat: Szerzők affiliációinak azonosítását követően előállt tábla példa	104
8. táblázat: Példa szerzői hálózatok éllistájára	105
9. táblázat: Példa affiliációkat is tartalmazó éllistára	106
10. táblázat: Példa hálózat éleinek térképi vonalobjektumként történő megadására	106
11. táblázat: Hasonlósági mátrixok összehasonlítása kapcsolaterősségi szintenként	109
12. táblázat: A különböző klaszterezési algoritmusok kofenetikus korrelációs együtthatói	111
13. táblázat: A hierarchikus klaszteranalízis legnagyobb tíz és öt klaszterének relatív mérete	114
14. táblázat: A szerzők klasztercsoportokba történő sorolásának hasonlósága módosított Rand-index alapján a közgazdaságtan területén	114
15. táblázat: A szerzők klasztercsoportokba történő sorolásának hasonlósága módosított Rand-index alapján a természetföldrajz területén	115
16. táblázat: Web of Science rekord Cited References (CR) mezője és abból a kinyert folyóirat	116
17. táblázat: Az öt legnagyobb szerzői klaszter öt leggyakoribb folyóirata és ez alapján a klaszter tématerület-elnevezései mindhárom távolságtípus esetében a közgazdaságtan területén	117
18. táblázat: Az öt legnagyobb szerzői klaszter öt leggyakoribb folyóirata és ez alapján a klaszter tématerület-elnevezései mindhárom távolságtípus esetében a természetföldrajz területén	117
19. táblázat: Hálózatok fő paraméterei	119
20. táblázat: Szerzők közötti átlagos hálózati távolságok Pearson-korrelációja, közgazdaságtan	125
21. táblázat: Szerzők közötti átlagos hálózati távolságok Pearson-korrelációja, természetföldrajz	125
22. táblázat: QAP korreláció eredménye a közgazdaságtan területén	127

23. táblázat: QAP korreláció eredménye a természetföldrajz területén	128
24. táblázat: A legnagyobb modulok szerzőinek aránya a teljes szerzőszámhoz viszonyítva a közgazdaságtan és a természetföldrajz területén	133
25. táblázat: A szerzők hálózati modularitás alapján történő csoportokba sorolásának hasonlósága módosított Rand-index alapján a közgazdaságtan területén	134
26. táblázat: A szerzők hálózati modularitás alapján történő csoportokba sorolásának hasonlósága módosított Rand-index alapján a természetföldrajz területén	134
27. táblázat: Az öt legnagyobb szerzői modul öt leggyakoribb folyóirata és ez alapján a modulok tématerület-elnevezései mindhárom távolságtípus esetében a közgazdaságtan területén	135
28. táblázat: Az öt legnagyobb szerzői modul öt leggyakoribb folyóirata és ez alapján a modulok tématerület-elnevezései mindhárom távolságtípus esetében a természetföldrajz területén	136

